International Symposium on Computer Science, Computer Engineering and Educational Technology (ISCSET-2021)

Prof. Dr. Dr. h. c. Uranchimeg Tudevdagva,
Prof. Dr. Dr. h. c. Wolfram Hardt (Hrsg.)

TUD*press*

IBS Scientific Workshop Proceedings
Herausgegeben von Stiftung IBS, Wolfram Hardt
Band 14

International Symposium on Computer Science, Computer Engineering and Educational Technology (ISCSET-2021)

Prof. Dr. Dr. h. c. Uranchimeg Tudevdagva,
Prof. Dr. Dr. h. c. Wolfram Hardt (Hrsg.)

TUD*press*
2021

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

ISBN 978-3-95908-238-9

THELEM Universitätsverlag
GmbH und Co. KG
http://www.tudpress.de

Gesetzt von den Herausgebern.
Printed in Germany.

IBS Scientific Workshop Proceedings

The "International Symposium on Computer Science, Computer Engineering and Educational Technologies" (ISCSET-2021) is the proceedings of the annual international symposium funded by the "Internationales Informatik- und Begegnungszentrum Sachsen (IBS)".

Continuing already for a decade the annual symposium is held successfully in cooperation of universities from Germany, Mongolia, Russia and China. This year university from Kazakhstan officially joined to our community.

To the ISCSET 2021 numerous contributions were submitted and all submitted papers went through review commissions which consists from professors and doctors of partner universities. Nearly 80% of the submission was accepted by reviewers and included to this proceeding.

Increasing number of submissions confirms success of our symposium and we are willing to upgrade our event to be more visible in community like IEEE.

We welcome universities, professors and scholars to our symposium and we believe that interdisciplinary research and study will support many good ideas in science and engineering education.

Prof. Dr. Dr. h. c. Uranchimeg Tudevdagva

Mongolian University of Science and Technology,
Chemnitz University of Technology

Prof. Dr. Dr. h. c. Wolfram Hardt

Chemnitz University of Technology

November 2021

Table of Contents

International Symposium on Computer Science, Computer Engineering and Educational Technology (ISCSET-2021)

The Impact of Distance Learning Quality on Student Satisfaction and Continuance Usage Intentions During COVID-19 5
Ariunaa Khashkhuu

Computational Tomographic Introscopy of Seismic-Acoustic Environments Under Conditions of Kinematic Linearization 12
Alexander Vokhmin, Nikolai Gervas, Sergey Zerkal, Egor Omelyanenko, Alexander Peshkov

Parametric Study for Lightweight Monocular Depth Estimation Deep Neural Network Model…………………………………………………………..…… 16
Abeer Aly Moussa, Shadi Saleh, Osama Mohamed El-Sayed, Abubaker Gaber, Batbayar Battseren, Wolfram Hardt

Deep-Learning-Based Insulator Detector for Edge Computing Platforms............ 21
Batbayar Battseren, Mohamed Salim Harras, Soaibuzzman

Technology and Access to Education During COVID-19.................................... 29
Bayartsetseg Batjav, Myagmarjav Usny-Ekh, Sanjaabadam Sed, Amarzaya Gulguu

Implementation of Distance Education Initiative Within Mongolian Higher Education Institutions During COVID-19 34
Buyandelger Banzragch, Lkhamtseren Bold, Gerelmaa Baljinnyam

Design of Two-Stream Convolutional Neural Network for Vision-Based Obstacle Avoidance in Indoor Mobile Robots.. 41
Chaymae El Mechal, Najiba El Amrani El Idrissi, Mostefa Mesbah, Chaymae Zai

Educational, Psychological and Behavioral Profiling With High Dimensionality Reduction... 47
Cyryl Koshyk

Research on Building Competition-Practice Platform and Promoting the Construction of Electronic-Technology Course.. 53
Hui Sun, Haibin Wu, Yang Cui, Honghao Wang, Xiaoming Sun, Dezhong Ma

Towards the Perception of Pedestrian Detection Using Vision Based Systems.... 57
Julkar Nine, Jayashree Mohan, Sweta Sharma, Shadi Saleh, Sanjay Balamurugna, Wolfram Hardt

Feature Points Detection and Matching in Lumen 3D Reconstruction Based on SLAM 63
Kaiyang Xu, Haibin Wu, Ying Wang, Jianbo Zhao, Yan Zhang, Ruotong Xu

Situation Analysis of ICT Incorporation Into Education and Teaching 70
Lkhagvasuren Purev, Buyandelger Banzragch, Nyamsuren Boloo

Cluster Analysis on Mathematical Ability of the Children.................................... 78
Narantuya Jamyandorj, Ganbat Tsend

Redefining Blockchain for IoT and Embedded Systems 82
Nidhee Kamble

The Study of the Information Culture Competencies of Secondary School Teachers...…………………………………………………………………………83
Nyamsuren Amsarvaa, Tsedendamba Sodov

Towards End-to-End Estimation of Camera Trajectory With Deep Monocular Visual Odometry 88
Sanketh Rajendra, Shadi Saleh, Batbayar Battseren, Julkar Nine, Wolfram Hardt

Toward CNNs Visualization for Estimation Depth From Monocular Camera 93
Shweta Dumbre, Shadi Saleh, Batbayar Battseren, Wolfram Hardt

Emotion Identification Research for Neural Network Training 99
Sofia Kharyutkina

Training Spatial Ability Using Geogebra Interactive Dynamic Software 101
Suvd Erdenechimeg, Batsuren Baasanjargal, Ganbat Danaa

Implementation of the Online Professional Development Program for Mass University Teachers During COVID-19 Pandemic 106
Tserenchimed Purevsuren, Tsooj Shambaljamts, Ariunbolor Davaa, Ganbat Danaa, Tumenbold Dashzeveg, Naidandorj Radnaa

The Evaluation of Online Yoga Course During Pandemic Period 113
Uranchimeg Tudevdagva, Zandi Nyamjav

The Evaluation of Adult Training Courses During the COVID-19 Quarantine.. 118
Uranchimeg Tudevdagva, Enkhtsetseg Luvsandagva, Altantsetseg Surenjav

Conversion of Traditional Course to Online Course- Empirical Evidence on Students Learning Experience.. 122
Ummay Ubaida Shegupta, René Schmidt, Wolfram Hardt

Lightweight Monocular Depth Estimation on Embedded Systems..................... 128
Vishnuvarthanan Srinivasan, Shadi Saleh, Batbayar Battseren, Wolfram Hardt

Training of the Masters' Students in the Smart City Program Using the Internet of Things Course... 135
Zhanar Akhayeva, Alma Zakirova, Aigul Nurpeissova

The Impact of Distance Learning Quality on Student Satisfaction and Continuance Usage Intentions During COVID-19

Ariunaa Khashkhuu
Department of Business Administration
School of Business Administration and Humanities
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
kh.ariunaa@must.edu.mn

Abstract— **Due to the COVID-19 pandemic, the Mongolian University of Science and Technology is in its third semester due to the distance learning. The study examines the factors influencing the behavioral intention of using technology in distance learning for formal educational purposes using the Technology Acceptance Model (TAM), as well as has also been expanded to include SERVQUAL dimensions, system quality and instructor materials quality dimensions. The TAM model identifies the impact of technological perceived usefulness and perceived ease of use on attitudes and behavioral intention. The ultimate goal of all these analyzes is to identify the factors that contribute to the satisfaction of students in distance learning. The questionnaire, which included the above criteria, was evaluated by 186 students, and the theoretical design sub-measures were grouped by quality criteria using Principal Component Analysis (PCA) and Structural Equation Modeling (SEM). In this study, the most important criteria for TAM are technological perceived usefulness, which affects the quality of teaching materials and the time flexibility and learning flexibility, as well as the impact of service quality and system quality on ease of use and the dimensions proved to be a sub-component of technology-based service quality. The Kruskal Wallis and Mann-Whitney U tests also confirmed that all sub-dimensions of this model differ in the extent of the impact of COVID-19, and that its three sub-dimensions differ in gender.**

Keywords— SERVQUAL, course materials quality, system quality, percieved usefulness and ease of use, behavioral intention and satisfaction

I. INTRODUCTION

The COVID-19 epidemic has caused the most significant disruption to education systems in human history, affecting almost 1.6 billion students in over 200 nations. More than 94 percent of the world's student population has been touched by school, institution, and other learning facility closures. This has resulted in significant changes in many parts of our life. Traditional educational procedures have been substantially disrupted by social distance and limited mobility regulations. Reopening schools once restrictions have been lifted is another problem, with numerous new standard operating procedures in place [1].

Due to the COVID-19 epidemic, most nations have implemented lockdown and social distancing measures, resulting in the shutdown of schools, training institutions, and higher education facilities. A fundamental shift is occurring in the way educators offer excellent education—via multiple online venues. Despite the obstacles that educators and learners face, online learning, remote learning, and continuing education have emerged as a cure for this unprecedented worldwide epidemic.

Transitioning from traditional face-to-face learning to online learning may be a completely different experience for both learners and educators, one they must adjust to with few or no other options. The school system and instructors have accepted “Education in Emergency” via different online platforms and are being forced to adopt a system for which they are unprepared.

During this epidemic, e-learning platforms played a critical role in assisting schools and universities in facilitating student learning when universities and schools were closed [2]. While adjusting to the new changes, staff and student preparedness must be assessed and supported as needed.

II. E-LEARNING CLASSIFICATIONS

TABLE I. E-LEARNING SYSTEMS CLASSIFICATIONS

Classification	Presence	eCommunication
Type I: Face-to-face	Physical	- post lecture notes - schedule assignments - discussion and e-mail outside classroom
Type II: Self-learner	None	None
Type III: Asynchronous	None	- includes all listed for Type I - audio/video lecture recordings
Type IV: Synchronous	Virtual	- includes all listed for Type III - “live” audio - “live” video - synchronous chat
Type V: Blended/Hybrid-Asynchronous	Physical	includes all listed for Type III
Type VI: Blended/Hybrid-Synchronous	Physical and Virtual	includes all listed for Type IV

The use of appropriate and relevant pedagogy for online education may be dependent on both instructors' and learners' knowledge and exposure to information and communications technology (ICT). Some of the online platforms that have been utilized so far include unified communication and collaboration platforms like Microsoft Teams, Google Classroom, Canvas, and Blackboard, which allow teachers to build educational courses, training, and skill development programs [3]. They offer office chat, video meetings, and file storage tools to keep classes organized and simple to work with. They often allow for the exchange of a wide range of material, including Word, PDF, Excel, audio, and video files, among others.

ICT-supported learning is rife with terminology and ideas that are used interchangeably or with minor variations to address the use of ICT in education: E-learning, Distributed Learning, Virtual Education, Internet-Based Education, Online Learning, Flexible Learning, Synchronous Learning, Web-Based Training, and so on are some examples. Negash, S., and M. V. Wilcox suggest six different types of e-learning categories [4]. Table I summarizes the six categories.

III. Technology Acceptance Model

The technology acceptance model (TAM) is one of the most commonly researched and utilized technology acceptance models. The two essential elements for a person's attitude (ATT) toward a utilized technology are perceived utility (PU) and perceived ease of use (PEU), which influences real system utilization. PU is a person's subjective perception that the use of a particular technology enhances individual job performance, whereas PEU represents a person's assessment of how much effort is required to use the new technology. Both variables are impacted by a variety of external factors, including work relevance, subjective norm, and output quality [5]. Davis et al. modified the model by introducing behavioral intention (BI) as a mediator between ATT and actual system use [6].

Lee et al. introduced perceived enjoyment (PE) as an intrinsic motivation to TAM components for the e-learning setting, in addition to PU and PEU [7]. Šumak et al. did a meta-analysis and discovered that the TAM was the most commonly used model in e-learning and that the extent of the causal effects between different TAM-related parameters varies depending on the user type and the kind of e-learning technology [8].

For our study, we expanded the research model of Davis and Venkatesh, as presented in Figure 1 [9].

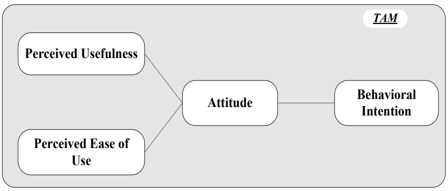


Fig. 1. The Technology Acceptance Model [5-6]

We consider that the TAM in e-learning is often studied in situations where blended learning or e-learning is utilized in addition to face-to-face teaching, but in the COVID-19 semester, virtual teaching and learning was the sole channel used to communicate material. We investigate the measurement model and the structural model before comparing the findings across time.

Acceptance Over Time

Venkatesh and Davis tested an extended TAM (TAM2) in four longitudinal studies and introduced experience as a relevant influencing factor that is important for understanding changes in PU over time, whereby experience in general reflects an opportunity to use a technology and is typically operationalized as the passage of time from an individual's first use of a technology [5]. Venkatesh et al. created UTAUT based on the TAM and tested it in a long-term field research [9]. Venkatesh et al. added three new dimensions to UTAUT, evaluated user experiences, and studied their impact on user acceptability and habits [10].

Davis and Wong used TAM2 in an educational setting to assess user experience in relation to real student usage (system use) of an e-learning system [11]. They emphasized the numerous underlying interactions that occur throughout e-learning adoption processes and suggested that future research use a longitudinal approach. Pynoo et al. investigated the acceptability of digital learning environments in the educational setting using UTAUT and discovered variations over time; they also said that the value of digital learning environments should be proven to optimize its utilization [12].

Unlike other research, ours focuses on the experience with technology-mediated education in the COVID-19 setting rather than a specific technology. We anticipate and demonstrate in this context that students will gain experience over the semester, resulting in quantifiable improvements in their acceptance.

Benefits and Disadvantages of E-Learning

In the aftermath of the COVID-19 crisis, the benefits and drawbacks of technology-mediated teaching and learning became a focus topic for university study [13-16] However, this is not a new issue, but it is one of the primary study areas in the context of learning in digital learning environments. E-learning, according to Davis and Wong, is a global phenomenon for companies and educational institutions that aims to improve students' learning experiences and effectiveness in terms of learning outcomes [11]. Recent study has explored the benefits of e-learning, but there is no agreement on whether the outputs of e-learning are more effective than those of traditional learning formats [17]. The most frequently mentioned advantages are cost efficiency, flexibility (in terms of time and place), saving time to travel to the learning location, easy access to learning materials, and the usefulness of learning materials for a longer period of time [18-24], or the ability to offer personalized learning based on the learner's specific needs [25].

On the negative side, technology-mediated learning lacks direct social connection and a personal touch, and thus has the ability to socially isolate the learner or, at the at least, negatively affect social elements of learning processes [20, 23-24, 26]. Socially isolated learning can have a negative impact on the development of learners' communication skills and change communication conditions, such as a lack of support and feedback through nonverbal cues or by observing the interactions of others, as well as a lack of social and cognitive presence and teacher involvement.

Furthermore, in the lack of regular touch with professors, students are uneasy about their learning [23]. Technology-mediated teaching and learning need self-motivation, time management, and a focused approach, as well as learners' self-directed learning and organizational abilities [20, 21]. According to Al-Qahtani and Higgins, these needs stem in part from social isolation and a lack of direct social connection, implying that the learner must be reasonably motivated to counteract this impact [23].

The evaluation models are developed for quality measurement of e-learning. The structure-oriented evaluation model (SURE) is one of the models which is focused on evaluation of e-learning [27]. The SURE model developed originally in 2014 by Tudevdagva as evaluation model foe e-learning and later extended to evaluation of complex systems [28]. Due to the pandemic e-learning is implemented for all level of teaching and many researchers doing self-evaluation based on the SURE model as well as by other models [29-31].

During the university lockout, the idea was that youngest students would have little trouble transitioning to online learning, which has been supported by actual findings [13]. Shah et al. highlight the numerous and obvious benefits of transferring learning to the virtual world: free exchange of information, access to lectures and presentations at conferences that previously required significant travel costs, webinars and online discussions, reduction of time inefficiency associated with travel, and increased commitment [16]. Owusu-Fordjour et al. highlight negative consequences, such as studying at home being unproductive due to numerous distractions, a lack of a suitable learning environment, or interaction with the teacher [15]. On the positive side, switching to online teaching has been found to have fewer problems; however, on the negative side, technical obstacles as well as a lack of communication and cooperation, difficulties concentrating, too much screen time, a lack of logistical infrastructure, non-physical presence, increased workload and the loss of lab courses, and a general restriction of social contact have been highlighted as important during the crisis. Positive features include ease of participation in class, time savings, home comfort, the ability to study, new competencies, attendance and learning flexibility, and so on.

IV. METHODS

The TAM model identifies the impact of technological perceived usefulness and perceived ease of use on attitudes and behavioral intention. The ultimate goal of all these analyzes is to identify the factors that contribute to the satisfaction of students in distance learning. The questionnaire, which included the above criteria, was evaluated by 186 students, and the theoretical design sub-measures were grouped by quality criteria using Principal Component Analysis (PCA) and Structural Equation Modeling (SEM).

The users who participated in the survey were asked to assess the importance of the 37 criteria and the seven sub-dimensions. The criteria were given as simply formulated sentences, so that the Likert five scale could be used. It should be noticed that the sample size was acceptable for factor analysis (the minimum sample size would have been 185, i.e. five times the number of variables).

TABLE II. PROPOSED CRITERIA FOR EACH OF THE SUB-DIMENSIONS OF DISTANCE LEARNING SERVICE QUALITY

Sub-dimensions	Criteria	Sub-dimensions	Criteria
Empathy	EMP1	Perceived ease of use	PEU1
	EMP2		PEU2
	EMP3		PEU3
	EMP4		
Assurance	ASS1	System quality	SysQ1
	ASS2		SysQ2
	ASS3		SysQ3
	ASS4		SysQ4
Responsiveness	RES1		SysQ5
	RES2		SysQ6
	RES3		SysQ7
	RES4	Flexibility	Flex1
Reliability	REL2		Flex2
	REL3		Flex3
Instructor materials quality	MQ1	Behavioral Intention	BI1
	MQ2		BI2
	MQ3		
	MQ4		

V. RESULTS

The demographic characteristics are (1) gender, (2) age, (3) location (UB city or rural) (4) department/field, and (5) How many years are you studying in Table III.

TABLE III. DEMOGRAPHIC PROFLE OF RESPONDENTS (N=186)

Gender	Frequency	Percent
Female	122	65.6%
Male	64	34.4%
Total	186	100.0%
Department		
Business Administration	139	74.7%
Technology management	6	3.2%
Social Science	3	1.6%
Humanities	9	4.8%
Other	29	15.6%
Total	186	100.0%
How many years are you studying?		
Freshman	9	4.8%
Sophomore	63	33.9%
Junior	53	28.5%
Senior	61	32.8%
Total	186	100.0%

Factor analysis results

It can be deduced that the 37 quality criteria were properly grouped into the seven sub-dimensions of the theoretical framework which was proposed to delineate the distance learning service quality construct. The KMO value was found to 0.939, that the sample size was adequate for applying factor analysis. Now I can proceed for Factor Analysis. 0.939 or 93.9% of the respondents provided excellent sample reliability. In the factor analysis of the above questions, the averages are similar, grouping into 7 sets. The results of the survey correctly explain 72.3% (Eigenvalues) of the total sample.

TABLE IV. KMO AND BARTLETT'S TEST

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.939
Bartlett's Test of Sphericity	4578.292	11354.016
	703	780
	.000	.000

The outcomes of structural equation modeling

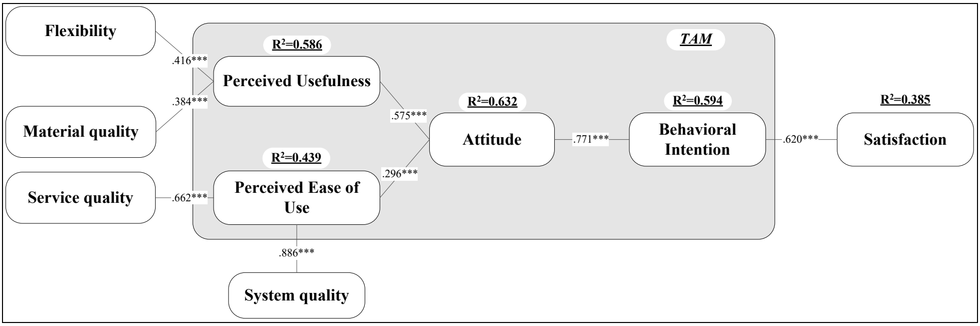
A good indicator for measuring a construct's consistency is Cronbach's alpha. A value above 0.7 indicates that the examined construct has consistency. Another valid indicator for measuring a construct's consistency is that factor loadings, also known as λ, should receive values more than 0.7, while the component reliability (CR) and the average variance extracted (AVE) index should exceed 0.8 and 50% respectively. In Table V, the proposed reliability measures along with each factor loadings are presented.

TABLE V. LOADING CALCULATIONS AND RELIABILITY MEASURES FOR EACH CONSTRUCT, INCLUDING COMPOSITE RELIABILITY (CR), AVERAGE VARIANCE EXTRACTED (AVE), AND CRONBACH'S ALPHA (A).

Variable	Criteria/Code	Loadings (λ)	Cronbach a	CR	AVE
Empathy			0.928	0.950	83%
	EMP1	.908			
	EMP2	.882			
	EMP3	.923			
	EMP4	.922			
Assurance			0.873	0.916	73%
	ASS1	.857			
	ASS2	.814			
	ASS3	.836			
	ASS4	.911			
Responsiveness			0.905	0.934	78%
	RES1	.901			
	RES2	.889			
	RES3	.895			
	RES4	.847			
Reliability			0.885	0.914	84%
	REL2	.926			
	REL3	.909			

Code	Dimension of service quality	Loadings (λ)	Cronbach a	CR	AVE
RES	Responsiveness	.965	0.946	0.962	86%
REL	Reliability	.943			
ASS	Assurance	.916			
EMP	Empathy	.896			

	Criteria	Loadings (λ)	Cronbach a	CR	AVE
Instructor materials quality			0.905	0.934	78%
	MQ1	.894			
	MQ2	.860			
	MQ3	.873			
	MQ4	.904			
Perceived usefulness			0.832	0.777	64%
	PU1	.808			
	PU2	.786			
Perceived ease of use			0.839	0.856	67%
	PEU1	.785			
	PEU2	.874			
	PEU3	.785			
System quality			0.917	0.935	67%
	SysQ1	.866			
	SysQ2	.861			
	SysQ3	.813			
	SysQ4	.877			
	SysQ5	.814			
	SysQ6	.771			
	SysQ7	.733			
	Flexibility		0.836	0.905	76%
	Flex1	.910			
	Flex2	.914			
	Flex3	.787			
Behavioral Intention			0.806	0.836	72%
	BI1	.836			
	BI2	.859			
Attitude		.908	.700	0.825	83%



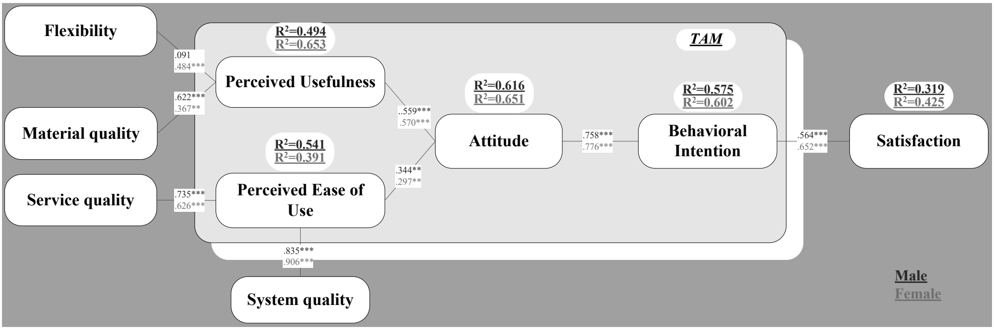
* p < 0.05 ** p < 0.01 *** p < 0.001

Fig. 2. Expended TAM test results

TABLE VI. MANN-WHITNEY U TEST STATISTIC, BY GENDER

	Empathy	Assurance	Responsiveness	Reliability	Instructor SQ	Flexibility	Perceived Ease of Use	Perceived Usefulness	Instructor MQ	System quality	Attitude	Behavioral Intention	Satisfaction
Mann-Whitney U	1642.500	1628.000	1510.500	1538.000	1559.000	1598.000	1415.500	1667.500	1350.500	1419.000	1672.500	1648.500	1701.500
Wilcoxon W	4963.500	4949.000	4831.500	4859.000	4880.000	4919.000	4736.500	4988.500	4671.500	4740.000	2618.500	2594.500	5022.500
Z	-.523	-.601	-1.219	-1.085	-.959	-.759	-1.729	-.394	-2.063	-1.699	-.377	-.494	-.217
Asymp. Sig. (2-tailed)	.601	.548	.223	.278	.338	.448	**.084**	.694	**.039**	**.089**	.706	.621	.828

a. Grouping Variable: Gender:



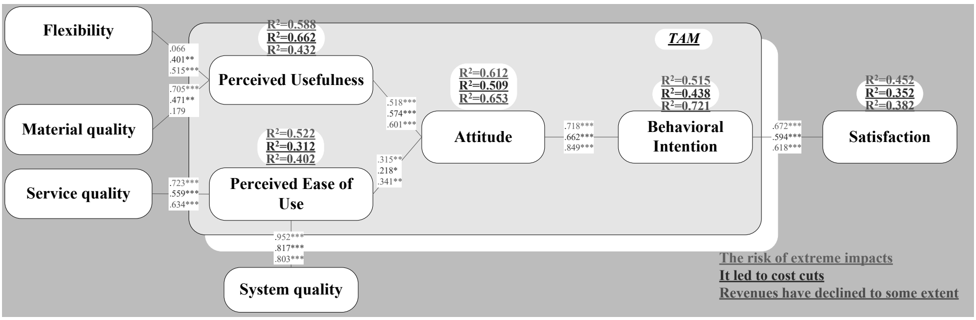
* p < 0.05 ** p < 0.01 *** p < 0.001

Fig. 3. Expended TAM test results by gender

TABLE VII. KRUSKAL WALLIS TEST STATISTIC, BY THE IMPACT OF COVID 19 ON YOU AND YOUR FAMILY'S FINANCIAL SITUATION

Test Statistics[a,b]													
	Empathy	Assurance	Responsiveness	Reliability	Instructor SQ	Flexibility	Perceived Ease of Use	Perceived Usefulness	Instructor MQ	System quality	Attitude	Behavioral Intention	Satisfaction
Chi-Square	19.025	11.755	11.424	10.338	14.122	18.956	11.911	14.066	14.150	11.679	20.227	12.389	11.464
df	2	2	2	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	.000	.003	.003	.006	.001	.000	.003	.001	.001	.003	.000	.002	.003

a. Kruskal Wallis Test
b. Grouping Variable: The impact of COVID 19 on you and your family's financial situation



* p < 0.05 ** p < 0.01 *** p < 0.001

Fig. 4. Expended TAM test results by by the impact of COVID 19 on you and your family's financial situation

Using the TAM, the blue background of Figure 2 shows the effects of factors influencing the behavioral goals of using the technologies used in distance learning for formal educational purposes. These include:

- ✓ An r-squared of 63.2% reveals that 63.2% of the data fit the regression model. Generally, a higher r-squared indicates a better fit for the model. 63.2% of the attitude variability (R^2) is explained by perceived usefulness (.575***), perceived ease of use (.296***), and perceived usefulness has a higher correlation (beta).
- ✓ 59.4% of behavioral intention variability are explained by student attitudes (0.771***), and the correlation between these two variable (beta) is positive.
- ✓ In the ANOVA table for the regression model, the F statistic is equal to 47.222/0.455 = 103.891. The distribution is F(2, 121), and the probability of observing a value greater than or equal to 103.891 is less than 0.001. Highly significant. There is strong evidence.

For expanded TAM:

1. The Instructor course Material Quality Dimensions (.384***) and the Training and Time Flexibility Measurement (.416***) account for 58.6% of variability in perceived usefulness. Its validity is confirmed by the ANOVA variance analysis with sign = 0.000. Highly significant. There is strong evidence. The beta coefficient shows that there is a positive correlation between the two sub-dimensions.
2. 43.9% of the variability in the ease of use of TAM are determined by the sub-dimensions of the quality of teaching services (determined by the instructor on the SERVQUAL dimensions, 0.662***). ANOVA, sign = 0.000. Highly significant. There is strong evidence. The beta coefficient shows that these sub-dimensions are strongly correlated.
3. The system quality sub-dimensions also has a very high correlation with ease of use (0.886). These two sub-dimensions were grouped into a single sub-dimensions by factor analysis.
4. The ultimate goal of all these analyzes is that 38.5 percent of the variability in satisfaction of students in distance learning is explained by behavioral goals, which is strongly correlated with beta = 0.620.

The model is differentiated by the impact of COVID-19 on your family's financial situation and by gender.

The Kruskal Wallis and Mann-Whitney U tests confirmed that all sub-dimensions of this model differ in the magnitude of the impact of COVID-19 on your family's financial situation and that the three sub-dimensions differ in gender. As shown in Table VI and Table VII. Therefore, the model is redefined in these two different ways, as shown in Figure 3 and Figure 4.

The Kruskal Wallis test tested three measures of COVID-19's impact on family's financial situation: "the risk of extreme impact", 'cost-cutting" and "revenues have declined to some extent", and confirmed that all sub-dimensions of the model were different. Based on this, the model was split into SPSS data for each of the three COVID-19 impact measures and the model dimensions were redefined. There are many differences in the impact of the model:

- ✓ "Instructor course material quality dimensions" (0.705***) explains 58.8% of the "perceived usefulness" variability in students who had a "strong" effect on the financial situation of the student's family, while a "slight" effect. For students, 43.2% of that variability is explained by flexibility (0.515) and the material quality (0.179), and in terms of correlation, flexibility explains most of this variability.

The gender based comparison of distance learning service quality sub-dimensions shown that there is three sub-dimensions difference between male and female with respect distance learning service quality sub-dimensions. These are: 1. Perceived ease of use, 2. Quality measure of instructor course materials, 3. System quality measure. There are many differences:

- ✓ For male students, 49.4% of the variability in "perceived usefulness" was explained by "instructor course material quality" (0.622***), and the correlation was strongly positive.
- ✓ For female students, 65.3% of the variability in "perceived usefulness" was explained by "flexibility" (0.484***) and "instructor course's material quality" (0.367***), and the correlation was positive for flexibility and the correlation was weakly positive for the material quality.

VI. Conclusions

Due to the COVID-19 pandemic, the Mongolian University of Science and Technology is in its third semester due to the distance learning. The study examines the factors influencing the behavioral intention of using technology in distance learning for formal educational purposes using the TAM.

The questionnaire, which included the above criteria, was evaluated by 186 students, and the theoretical design sub-measures were grouped by quality criteria using PCA and SEM. In this study, the most important criteria for TAM are technological perceived usefulness, which affects the quality of teaching materials and the time flexibility and learning flexibility, as well as the impact of service quality and system quality on ease of use and the dimensions proved to be a sub-component of technology-based service quality. The Kruskal Wallis and Mann-Whitney U tests also confirmed that all sub-dimensions of this model differ in the extent of the impact of COVID-19, and that its three sub-dimension difference between male and female.

VII. Future Research

Reduce the probability of error of the sample to 5%, therefore, the survey should be conducted on 384 students.

The model can be further expanded. For example, it is possible to include a measure of social isolation.

For students of different majors, the TAM model analysis provides interesting results.

It is also possible to evaluate a particular technology by the TAM model alone. For example, in the academic year of 2020-21, the Mongolian University of Science and Technology is fully using MsTeams software in its training. At the end of this semester, an evaluation of the TAM model will yield interesting results.

Finally, the TAM model can be tested at different cycles and times in terms of the time the technology has been in use and compared with previous estimates.

References

[1] Sumitra Pokhrel and Roshan Chhetri, "A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning," 2021. DOI: 10.1177/2347631120983481 journals.sagepub.com/home/hef

[2] Subedi, S., Nayaju, S., Subedi, S., Shah, S. K., & Shah, J. M., "Impact of e-learning during COVID-19 pandemic among nursing students and teachers of Nepal," International Journal of Science and Healthcare Research, 2020, 5(3), 9.

[3] Petrie, C, "Spotlight: Quality education for all during COVID-19 crisis," (hun1drED Research Report #01). United Nations, 2020. https://hundred.org/en/collections/quality[1]education-for-all-during-coronavirus

[4] Negash, S., & Wilcox, M. V, "E-learning classifications: Differences and similarities," In S. Negash, M. E. Whitman, A. B. Woszczynski, K. Hoganson & H. Mattord (Eds.), Handbook of distance learning for real-time and asynchronous information technology education 2008, (pp. 1-23). London: Information Science Reference

[5] Venkatesh, V., and Davis, F. D. "A theoretical extension of the technology acceptance model: four longitudinal field studies," Manage. Sci. 46, 2000, 186–204. doi: 10.1287/mnsc.46.2.186.11926

[6] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R., "User acceptance of computer technology: A comparison of two theoretical models," Management Science, 35(8), 1989, 982-1003.

[7] Lee, M. K., Cheung, C. M., and Chen, Z., "Acceptance of internet-based learning medium: the role of extrinsic and intrinsic motivation," Inform. Manage. 42, 2005, 1095–1104. doi: 10.1016/j.im.2003.10.007

[8] Šumak, B., Hericko, M., and Pušnik, M., "A meta-analysis of e-learning ˇtechnology acceptance: the role of user types and e-learning technology types," Comput. Hum. Behav. 27, 2011, 2067–2077. doi: 10.1016/j.chb.2011.08.005

[9] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D., "User acceptance of information technology: toward a unified view," MIS Q. 27, 2003, 425–478. doi: 10.2307/30036540

[10] Venkatesh, V., Thong, J. Y., and Xu, X. "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology," MIS Q. 36, 2012, 157–178. doi: 10.2307/4141 0412

[11] Davis, R., and Wong, D., "Conceptualizing and measuring the optimal experience of the eLearning environment". Decis. Sci. J. Innovat. Educ. 5, 2007, 97–126. doi: 10.1111/j.1540-4609.2007.00129.x

[12] Pynoo, B., Devolder, P., Tondeur, J., van Braak, J., Duyck, W., and Duyck, P., "Predicting secondary school teachers' acceptance and use of a digital learning environment: a cross-sectional study," Comput. Hum. Behav. 27, 2011, 568–575. doi: 10.1016/j.chb.2010.10.005

[13] Kamarianos, I., Adamopoulou, A., Lambropoulos, H., and Stamelos, G., "Towards an understanding of university students' response in times of pandemic crisis (COVID-19)," Eur. J. Educ. Stud. 7:7, 2020, doi: 10.46827/ejes.v7i7. 3149

[14] Karalis, T., and Raikou, N., "Teaching at the times of COVID-19: inferences and implications for higher education pedagogy," Int. J. Acad. Res. Bus. Soc. Sci. 10:5., 2020, doi: 10.6007/IJARBSS/v10-i5/7219

[15] Owusu-Fordjour, C., Koomson, C. K., and Hanson, D., "The impact of COVID-19 on learning-the perspective of the Ghanaian student," Eur. J. Educ. Stud. 7, 2020, 88–101.

[16] Shah, S., Diwan, S., Kohan, L., Rosenblum, D., Gharibo, Ch, Soin, A., et al., "The technological impact of COVID-19 on the future of education and health care delivery," Pain Physician 23, 2020, S367–S380.

[17] Derouin, R. E., Fritzsche, B. A., and Salas, E., "E-learning in organizations," J. Manage. 31, 2005, 920–940. doi: 10.1177/0149206305279815

[18] Welsh, E. T., Wanberg, C. R., Brown, K. G., and Simmering, M. J., E[1]learning: emerging uses, empirical results and future directions. Int. J. Train. Dev. 7, 2003, 245–258. doi: 10.1046/j.1360-3736.2003.00184.x

[19] Brown, L., Murphy, E., and Wade, V., "Corporate eLearning: human resource development implications for large and small organizations," Hum. Resour. Dev. Int. 9, 2006, 415–427. doi: 10.1080/13678860600893607

[20] Hameed, S., Badii, A., and Cullen, A. J., "Effective e-learning integration with traditional learning in a blended learning environment," in Proceedings of the European and Mediterranean Conference on Information Systems, (Dubai), 2008, 25–26.

[21] Jefferson, R. N., and Arnold, L. W., "Effects of virtual education on academic culture: perceived advantages and disadvantages," Online Subm. 6, 2009, 61–66.

[22] Hill, N. S., and Wouters, K., "Comparing apples and oranges: toward a typology for assessing e-learning effectiveness," Res. Pers. Hum. Resour. Manage. 29:201., 2010, doi: 10.1108/s0742-7301(2010)0000029008

[23] Al-Qahtani, A. A., and Higgins, S. E., "Effects of traditional, blended and e-learning on students' achievement in higher education," J. Comput. Assist. Learn. 29, 2013, 220–234. doi: 10.1111/j.1365-2729.2012.00490.x

[24] Becker, K., Newton, C., and Sawang, S., "A learner perspective on barriers to e-learning," Austr. J. Adult Learn. 53:211, 2013.

[25] Berge, Z. L., and Giles, L., "Implementing and sustaining e-learning in the workplace," Int. J. Inform. Commun. Technol. Educ. (IJICTE) 2, 2006, 64–75. doi: 10.4018/jicte.2006100106

[26] Gimson, A., and Bell, J., "E-learning: your flexible development friend?" Dev. Learn. Organ. Int. J. 21, 2007, 7–9. doi: 10.1108/14777280710828558

[27] Uranchimeg Tudevdagva: Structure Oriented Evaluation Model for E-Learning. Wissenschaftliche Schriftenreihe Eingebettete Selbstorganisierende Systeme, Universitätsverlag Chemnitz, Chemnitz, Germany, July 2014. 123 p., ISBN: 978-3-944640-20-4, ISSN: 2196-3932.

[28] Uranchimeg Tudevdagva, "Structure-Oriented Evaluation an Evaluation Approach for ¬Complex Processes and Systems", Gewerbestrasse 11, 6330 Cham, Switzerland, Springer, 2020, pp.92, ISBN 978-3-030-44805-9 ISBN 978-3-030-44806-6 (eBook), https://doi.org/10.1007/978-3-030-44806-6.

[29] Uranchimeg Tudevdagva, Tsooj Shambaljamts, Delgermaa Senden, "The Evaluation of Distance Teaching During COVID Quarantine at MUST," in Proceedings of International Symposium on Computer Science, Computer Engineering and Educational Technologies, ISCSET 2020, pp.90-93, Laubusch, Germany, ISBN 978-3-95908-223-5.

[30] Uranchimeg Tudevdagva, Bazarragchaa Sodnom, and Selenge Erdenechimeg, "The Evaluation Case Study of Online Course During Pandemic Period in Mongolia," in journal Computer Science & Information Technology (CS & IT), Vol.11, No.6, pp.29-40, David C. Wyld et al. (Eds): CONEDU, NBIoT, MLCL, ISPR-2021, International Conference of Education (CONEDU 2021), May 22~23, 2021, Zurich, Switzerland, ISBN : 978-1-925953-40-4

[31] Uranchimeg Tudevdagva, Ariane Heller, Wolfram Hardt, "The Evaluation Report of the Online Training for CUT and UJEP Students during Pandemic Period," In Proceedings of 10th International Conference on Educational and Information Technology, ICEIT 2021. pp.66-69, IEEE, Chengdu, China, DOI: 10.1109/ICEIT51700.2021.9375565

Computational Tomographic Introscopy of Seismic-Acoustic Environments Under Conditions of Kinematic Linearization

Alexander Vokhmin
Computer Engineering department
Novosibirsk State Technical University
Novosibirsk, Russia
alexandervokhmin@gmail.com

Nikolai Gervas
Computer Engineering department
Novosibirsk State Technical University
Novosibirsk, Russia
nik.gervas@mail.ru

Sergey Zerkal
Computer Engineering department
Novosibirsk State Technical University
Novosibirsk, Russia
zerkal@ngs.ru

Egor Omelyanenko
Computer Engineering department
Novosibirsk State Technical University
Novosibirsk, Russia
egaomel@gmail.com

Alexander Peshkov
Computer Engineering department
Novosibirsk State Technical University
Novosibirsk, Russia
mupeskov1997@mail.ru

***Abstract* — When processing, interpreting and presenting information from almost any experimental study, its integral component is the corresponding information-measuring system. When performing work related to industrial "Introscopy", kinematic diagnostics of the ocean, the Earth's interior and the atmosphere under the conditions of sounding by X-ray, acoustic and seismic signals, such a system is "Computed tomography" with its own information, computing and measuring base. The paper presents an approach to solving inverse tomographic problems arising in the field of acoustics, seismics, at the junction with the problems of local (ROI-tomography) industrial flaw detection (introscopy). An illustration of the possibilities of this approach is given in the form of a model computational computer experiment.**

***Keywords* — computed tomography, inverse problem, kinematic diagnostics, linearization, local tomographic ROI reconstruction.**

I. Introduction

Natural science problems that admit mathematical formulations and use the apparatus of applied mathematics to solve them can be divided into two closely related classes, namely, direct and inverse problems. The first, i.e. direct problems are characterized by the development of processes from cause to effect, while the solution of inverse problems consists in reversing cause-and-effect relationships, i.e. the known result of solving the direct problem (consequence) determines the cause that gave rise to the effect. An important section of inverse problems is "Inverse problems of mathematical physics and analysis". Computational introscopy and its mathematical models belong to this section and are considered in this article from the standpoint of Computer (computational) tomography, a relatively young scientific direction in applied mathematics. The effectiveness of the computed tomographic approach to the study of the object of research lies in the used information-measuring system, which provides, in the classical version, useful information in the fullest possible extent. This effect is achieved by "shining through" the object under study with working radiation from all possible directions (scanning angles in the range from 0 to 2π). The technical implementation of this tomographic measuring information system is medical and industrial tomographs (according to the field of application), which use X-ray radiation as a probing signal. It is important to note that it is obligatory here to have both a direct path for performing measurements using a tomographic technique and information and computing support necessary for computer processing and interpretation of measured data, which are called "projection" data [1, 2].

II. Tomographic Approach to Kinematic Introscopy under Conditions of Linearization of the Probing Signal Speed

So, according to the field of use, medical, industrial (introscopy), geophysical (geotomography), ATO (Acoustic Tomography of the Ocean) and other tomography are distinguished. By the used working radiation: X-ray, optical, seismoacoustic tomography, NMR (Nuclear Magnetic Resonance). In addition, tomography can be "transmission" (rays from the outside pass through the object under study) and "emission" (when the recorded signal is generated by the object under investigation). All of the above, in essence, refers to the scientific and practical concept of "Introscopy", defined as, translated from Greek, as "intravision". In the "Earth Sciences", by analogy, there is the term "seismic imaging" i.e. remote study of the earth's interior using a sounding seismic signal.

Currently, the term "Seismic tomography" is widely used in the scientific literature. However, since the specificity of the study does not allow the implementation of traditional formulations of tomographic problems and approaches to their solution, the studies considered, in this regard, refer to the so-called non-classical computed tomography in terms of local tomographic reconstruction, which in the English version is called ROI-tomography (Region of Interest). It should be noted that in this case we are talking about kinematic seismotomography using the "geometric seismic" approximation, by analogy with "geometric optics", i.e. the consideration includes ray trajectories and travel times of the sounding signal (refracted waves) along these trajectories. The definition is subject to the velocity distribution in the studied environment, which is directly related to its density characteristics. The main mathematical relationship (equation) is the Fermat functional, which connects the above-mentioned kinematic characteristics. In fact, the solution to the problem is to "invert" the integral transformation involved in the Fermat functional. This problem is known as the "Inverse kinematic seismic problem", here, in a three-dimensional formulation, ray trajectories are spatial curved lines, in contrast to straight lines in X-ray tomography, and the use of the computed tomography method, at first glance, is not possible. However, if we consider the arising problem for environment that allow linearization of the propagation velocity of refracted waves near a linear function of depth, then we obtain a situation that reduces to classical tomography, while the initial rays are replaced by arcs of circles (rays corresponding to a linear function of velocity), followed by a projection these arcs into chords located on the day surface. This procedure reduces the problem to the required rectilinear tomographic rays, and it becomes possible to use algorithms based both on the integral inversion formula and on the algebraization of the original integral transformation (algebraic reconstruction). It is important that the organization of the measurement system in the form of a circle located (for convenience) with the center at the origin of the coordinate system, i.e. at the point (0, 0, 0), allows you to remove the overdetermination of the problem. From the measured times of arrival of refracted waves, a function of two polar angles at the source and receiver of the signal, and the radius of the measurement system, a three-dimensional function is determined - the propagation velocity of the noted waves. In the previously investigated formulations, the time function depended on four variables $(x^0, y^0, 0)$ and $(x^1, y^1, 0)$ coordinates of the source and receiver locations, i.e. (x^0, y^0, x^1, y^1). As noted above, in the conditions of linearization, it becomes possible to replace real rays connecting the source and receiver with circular arcs (rays corresponding to the linear dependence of velocity on depth), based on the circumference of the measuring system and forming the surface of the "spherical segment" at the points of which the desired velocity is determined. By changing the radius of the circle of the measuring system, it is possible to construct a sequence of nested similar surfaces and obtain a three-dimensional velocity distribution within the specified spherical segment. In this case, in fact, tomographic reconstruction is performed on the plane $z = 0$, in a circle bounded by the circumference of the measurement system $x^2 + y^2 = R_0^2$. The above-mentioned chords connecting the respective receivers are used as tomographic beams. Recalculation of the values of the solution found as a result of tomographic reconstruction on the z = 0 plane on a second-order surface stretched over the circumference of the measurement system gives the desired velocity distribution. Note [3, 4] that algebraic reconstruction is promising for solving problems of non-classical tomography.

III. RESULTS

To illustrate the constructiveness of the above tomographic approach to kinematic introscopy under conditions of linearization of the probe signal velocity, let us consider a computational computer model experiment. Let the speed in the environment be determined by the refractive index:

$$n(z) = (1 + 0{,}5z)^{-1} + 0{,}1z^2, \qquad (1)$$

where $n(z)$ is the refractive index, the reciprocal of the velocity or "slowness".

Here, the first term is considered to be known and dominant, and the second, $n_1(z) = 0{,}1z^2$, is small compared to the first, and must be determined as a result of solving the inverse problem. The choice of such a velocity dependence (one-dimensional) does not limit the generality of the approach and is aimed at facilitating the solution of the direct problem (since the situation under consideration refers to multi-angle tomography with a large number of rays), i.e. calculation of the vector of changes, travel times of the refracted signal. For more complex velocity distributions, involving three variables, the direct problem involves solving a system of differential ray equations, with initial-boundary conditions. This case is complicated by an inclusion that is opaque for the probe radiation, which needs to be diagnosed against the background of reconstruction of the velocity distribution. Thus, we get a model that fits the definition of ROI-tomography. The opaque area is a vertical cylinder with a section radius ρ = 0.3 of the observation system radius $R_0 = 1$ (Fig. 1). Fig. 2 shows an image of the exact solution, and Fig. 3 shows the result of solving the inverse problem, the number of source-receiver pairs is 16 and the number of projections is 32^2 (the total number of rays 16384) is the reconstruction grid step, which determines the number of pixels (unknown in SLAE), was chosen so that at least three rays pass through the pixel. A parallel scheme for collecting projection data is used. Under the condition of moderate mathematical incorrectness of the problem, it is visually clear that the results are quite satisfactory. The opacity region was reconstructed with good spatial "resolution"; the error in reconstructing the velocity distribution has a grid error rate of the order of 10%. When solving the emerging SLAE, an iterative algorithm for sequential calculation of strings was used, in accordance with the mathematical method of computer tomography. Regularization was carried out by an iterative-descriptive combination of a moving average filter and median filtering [5]. Taking into account a priori information about the smoothness of the measurement vector (elements of the projection matrix) and the desired solution justifies the use of smoothing, for example, by bicubic splines. Note that in this case, smoothing was not applied in Fig. 2 there are intrinsic noises of the algorithm.

In the studied area of scientific interests there is an important problem of diagnostics of velocity distribution in heterogeneous environments, i.e. with a change in velocity (and density) characteristics over depth (variable $z, (x, y, z) \epsilon R^3$). The speed linearization processes are considered possible near the constant:

$$V(z) = A + \alpha(z), |\alpha(z)| << A \qquad (2)$$

Such formulations arise in acoustic studies of the ocean and the atmosphere, and are also effective in a number of cases related to the Earth. In this case, both dynamic (the wave equation acts as a mathematical model) and kinematic formulations, and the response of the environment to the probing action in the form of scattered to inhomogeneous signals is used [6]. Here, with the kinematic approach, both refracted waves from the source and those scattered from inhomogeneities (secondary sources) are involved; it is essential that both sources and receivers have narrow directional patterns. The dynamic approach uses an observing system in the form of a co-located source and receiver or spaced in depth. It should be noted that both approaches make it possible to diagnose waveguide phenomena, moreover, computational algorithms based on explicit formulas for the inversion of integral transformations for the desired velocity, discretization of the problem, and finite-difference regularization.

These studies are of interest for remote study of elastic environments in geophysics, ocean acoustics, acoustic and optical studies of the atmosphere, as well as defectoscopy, incl. and monitoring of foundations of large building structures. It is worth noting as possible applications metallurgy and woodworking industry. Moreover, when working with objects of metallurgy, diagnostics of temperature distribution, the working radiation is the thermal emanation of the object of research itself, which already refers to emission tomography. Under the conditions of industrial flaw detection, modern ROI-tomographic technologies make it possible to confidently diagnose "difficult" defects such as cracks and delamination, which have a selective sensitivity to probe radiation [7, 8, 9, 10].

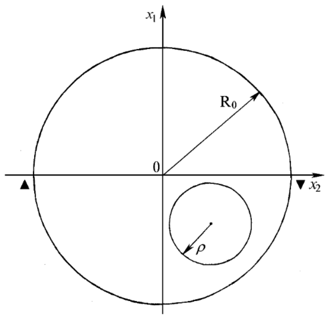


Fig. 1. Illustration of the measurement system and opacity locations R_0 - circle radius, signal sources and receivers locations, ▲ - source, ▼ - receiver, ρ - radius of the cylinder section (opacity area).

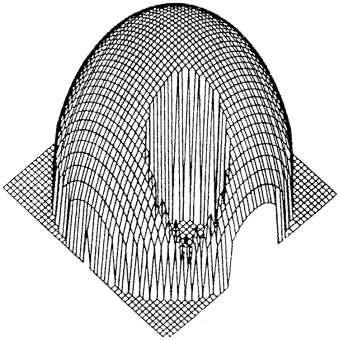


Fig. 2. Exact solution with the largest value of $n_1(z)$ within the study area is $5{,}76 * 10^{-3}$.

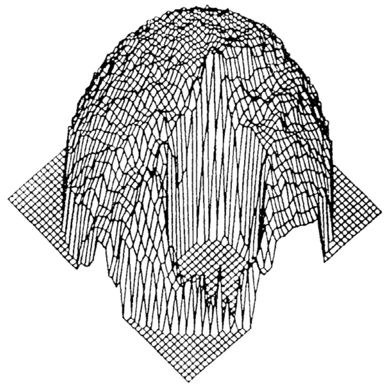


Fig. 3. The result of solving the inverse problem with the largest value $5{,}27 * 10^{-3}$.

For clarity, Fig. 2 and Fig. 3 are offset along the cylinder axis. The number of pixels (elements of the tomographic reconstruction grid) is 2500 elements.

The development of modern information-measuring systems ensures the continuity of tomographic techniques developed in various fields of knowledge. In particular, this refers to the methods of replenishing projection data in various situations of their incompleteness. The studies carried out in [11, 12] seem to be promising for the post-processing of tomograms (images obtained as a result of tomographic reconstruction). In addition, for further research, it is of interest to generalize the tomography in the cone of rays [13] to the tomographic approach presented in this work.

IV. CONCLUSION

The innovative approach outlined in the article, based on the capabilities of modern computer technology of information technology, allowing to process volumes of information that were not even considered in the second half of the last century during the rapid development of computed tomography, is aimed at solving complex unconventional tomographic problems. Algebraic reconstruction from the point of view of local tomography has wide opportunities for introscopy problems that are applicable both in industrial X-ray flaw detection, and good prospects for seismic-acoustic studies under conditions of restrictions on projection data.

REFERENCES

[1] V. V. Pikalov, "Application package for computational tomography [Paket prikladnykh programm, oriyentirovannyy na zadachi vychislitel'noy tomografii] (in Russian)," Reconstructive tomography issues [Voprosy rekonstruktivnoy tomografii], Novosibirsk, 1985, pp. 132–135.

[2] V. V. Pikalov and N. G. Preobrazhensky, "Reconstructive tomography in gas dynamics and plasma physics [Rekonstruktivnaya tomografiya v gazovoy dinamike i fizike plazmy] (in Russian)," Science [Nauka], Novosibirsk, 1987, p. 230.

[3] O. N. Belousova and S. M. Zerkal, "Numerical study of the algorithm for choosing a reference velocity distribution in the problem of kinematic seismotomography [Chislennoye issledovaniye algoritma vybora referentnogo skorostnogo raspredeleniya v zadache

kinematicheskoy seysmotomografii] (in Russian)," Autometry [Avtometryya], 2014, №6, pp. 120–127.

[4] S. M. Zerkal, "Tomographic diagnostics of velocity distribution in seismic environments [Tomograficheskaya diagnostika skorostnogo raspredeleniya v seysmicheskikh sredakh] (in Russian)," Autometry [Avtometryya], 2008, vol. 44, №1, pp. 82–87.

[5] A. V. Bronnikov and Y. E. Voskoboinikov, "Combined algorithms for nonlinear filtering of protected signals and images [Kombinirovannyye algoritmy nelineynoy fil'tratsii zashchishchennykh signalov i izobrazheniy] (in Russian)," Autometry [Avtometryya], 1990, №1, pp. 56–66.

[6] V. I. Dobrinsky and M. M. Lavrentiev, "Method for determining the profile of the speed of sound by recording the acoustic field at one point [Sposob opredeleniya profilya skorosti zvuka po zapisi akusticheskogo polya v odnoy tochke] (in Russian)," Acoustic wave propagation [Rasprostraneniye akusticheskikh voln], Vladivostok, 1978, pp. 12–14.

[7] S. M. Zerkal, "Computer solution for one class of flaw detection problems [Resheniye na EVM odnogo klassa zadach defektoskopii] (in Russian)," Report of the USSR Academy of Sciences [Doklad AN SSSR], 1990, vol. 313, №1, pp. 180–182.

[8] S. M. Zerkal, "Local tomographic reconstruction of the envelope of a thin defect using a reference sample [Lokal'naya tomograficheskaya rekonstruktsiya ogibayushchey tonkogo defekta s ispol'zovaniyem etalonnogo obraztsa] (in Russian)," Siberian Journal of Industrial Mathematics [Sibirskiy zhurnal industrial'noy matematiki], Novosibirsk, 2000, vol. 3, №1(5), pp. 110–115.

[9] R. S. Abashin, A. V. Likhachev, S. M. Zerkal, N. A. Likhachev, "Comparison of methods for processing of X-ray images of defects in reinforced concrete products," IOP Conference Series: Materials Science and Engineering, 2020, vol. 953 : 13 International Scientific Conference Architecture and Construction, Novosibirsk, 22-24 Sept. 2020. - Art. 012041 (8 p.). - DOI: 10.1088/1757-899X/953/1/012041.

[10] A. V. Likhachev, S. M. Zerkal, N. A. Likhachev, "Application of wavelet analysis for processing tomograms of narrow cracks," IOP Conference Series: Materials Science and Engineering, 2020, vol. 953 : 13 International Scientific Conference Architecture and Construction, Novosibirsk, 22-24 Sept. 2020. - Art. 012040 (8 p.). - DOI: 10.1088/1757-899X/953/1/012040.

[11] Y. E. Voskoboinikov and V. A. Boeva, "Robust nonparametric identification algorithm [Ustoychivyy algoritm neparametricheskoy identifikatsii] (in Russian)," Computing technology [Vychislitel'nyye tekhnologii], 2020, vol. 50, №3, pp. 46–53.

[12] Y. E. Voskoboinikov, "Wavelet filtering artifacts of images and their elimination," Optoelectronics, Instrumentation and Date Processing 2020, vol. 56, issue 6.

[13] O. N. Belousova, S. M. Zerkal, E. V. Shaposhnikova, "Innovative beam-cone computed tomography [Innovatsionnaya vychislitel'naya tomografiya v konuse luchey] (in Russian)," Modeling and high-tech information technologies in technical and socio-economic systems: Tr. 5 int. scientific-practical conference [Modelirovaniye i naukoyemkiye informatsionnyye tekhnologii v tekhnicheskikh i sotsial'no-ekonomicheskikh sistemakh : tr. 5 mezhdunar. nauch.-prakt. konferentsiya], Novokuznetsk, 2021, pp. 103–108.

[14] N.V. Vazhentseva and S. M. Zerkal, "Mathematical modeling in the tasks of industrial tomographic diagnostics in conditions of incomplete projection data [Matematicheskoye modelirovaniye v zadachakh promyshlennoy tomograficheskoy daignostiki v usloviyakh nepolnoty proyektsionnykh dannykh] (in Russian)," Proceedings of universities. Ferrous metallurgy [Izvestiya vuzov. Chernaya metallurgiya], 2013, №10, pp. 43–46.

Parametric Study for Lightweight Monocular Depth Estimation Deep Neural Network Model

Abeer Aly Moussa
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
abeer-aly-abdelaz.aly-moussa@s2018.tu-chemnitz.de

Shadi Saleh
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
shadi.saleh@informatik.tu-chemnitz.de

Osama Mohamed El-Sayed
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
osama.abdelaal@s2020.tu-chemnitz.de

Abubaker Gaber
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
abubaker-gaber-ahmed.gaber@s2020.tu-chemnitz.de

Batbayar Battseren
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
batbayar.battseren@informatik.tu-chemnitz.de

Wolfram Hardt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
wolfram.hardt@informatik.tu-chemnitz.de

***Abstract*— With the advance in autonomous vehicles, the need to analyze the images of road objects increased. That motivated many researchers to develop a state-of-art deep learning model to estimate the depth from the captured images. Currently, accurate depth estimation results from images are a fundamental portion of many applications, particularly in self-driving cars. Due to the high volume of training data, the cost of model training is very high. Lead to complexity to found the best hyperparameters used for training procedures. This paper review the previous state-of-art monocular depth estimation models and the most recent dataset that can fill the requirement to train a model. Besides, a set of experiments was conducted to optimize selected hyperparameters of a monocular depth estimation model. That led to finding the best hyperparameters with low computation costs. The evaluation metrics showed the best training accuracy at a learning rate equal to 1e-5 and Loss weight equal 0.55.**

***Keywords*— depth estimation; model hyperparameters; monocular; Autoencoder.**

I. INTRODUCTION

Depth estimation is a fundamental task in computer vision [1]. The task demands an input RGB image and outputs a depth image. Depth image includes statistics approximately the distance of the objects in the image from the viewpoint, usually the camera taking the image. It has numerous applications, such as self-driving cars, 3D scene reconstruction in robotics tasks, including localization, mapping, and obstacle avoidance. Convolutional neural networks (CNN) are deep neural network designs for processing a structured array of data. Computers see an input image as array pixels, and it depends on image resolution. array of matrix is (h = Height, w = Width, d = Dimension) h x w x d. such as an image of 6 x 6 x 3 array of RGB matrix (3 refers to RGB values).

Technically, estimating depth maps from single images can be considered a high-dimensional complex function where the input is an image. The output is the corresponding depth map [2]. The human brain has an astonishing capacity to understand visual scenes and solve this function by taking advantage of monocular cues in the image such as vanishing points, lighting, shading, perspective, and proportional scaling [3]. Monocular cues, experiences learned, knowledge about the world and the volume of familiar objects, their processing in the brain, and a 3D visualization of the scene obtained. It is an easy task for humans; even people with monocular (only one eye) can do it. However, implementing a system capable of using monocular for depth estimation is challenging [4].

Depth estimation is a heavy prediction task in which each pixel in the image is assigned a depth value. It is closely related to the semantic segmentation task, aiming to customize a class label for each pixel from several categories [5]. It is desirable to have an output prediction with the same input spatial size for dense prediction tasks. Each pixel has a matching prediction input achieved by using an Encoder-Decoder network architecture. Generally, the Encoder extracts valuable features from the input image and down-samples for featured maps. The Decoder then up-samples the feature maps, recover image resolution and produce pixel-wise predictions [6].

II. RESEARCH QUESTIONS

The aims of this paper can be broken down into three questions as follows:

- What kind of neural network architectures has previously been used to estimate depth from monocular images? and what makes them suited for this particular task?
- How does the proposed network perform when the training parameters change compared to similar state-of-the-art methods?
- How do the implemented network architectures, assuming training on the same dataset, differ in training time and performance?

III. RELATED WORK

There was a significant improvement in learning the methods of depth estimating over the past couple of years [7, 8, 9]. The majority of existing methods include deep learning network CNN trained on RGB images and depth maps corresponding. These methods can classify into supervised, semi-supervised, and self-supervised. Table I shows the most popular architectures and what is the category that is belonging to them.

A. Supervised monocular depth estimation.

Supervised learning is a process for artificial intelligence (A.I.), where a computer algorithm takes trained on input data classified for a specific output.

TABLE I. CATEGORIES OF DEEP LEARNING-BASED MONOCULAR DEPTH ESTIMATION ALGORITHM

Method	Architecture	Category
EMDEOM	FC	Supervised
ACAN	Encoder-Decoder	
Dense Depth	Encoder-Decoder	
DORN	CNN	
VNL	Encoder-Decoder	
BTS	Encoder-Decoder	
Deep V2D	CNN	
LISM	Encoder-Decoder	Self-Supervised
monoResMatch	CNN	
VOMonoDepth	Auto-Decoder	
GASDA	CNN	Semi-Supervised

FC – Fully Convolutional; CNN – Convolutional Neural Networks

The model introduced untottering the implicit patterns and relationships between the input data and the output labels, enabling it to produce accurate labelling results when presented with never-before-seen data.

Supervised methods take a single image and the corresponding depth information for training. In such a case, the network can directly train output Depth information. However, there is a need for many depth data of high quality, which are difficult to generalize to all use cases.

Cao et al. [10] proposed a convolutional Neural Network (CNN) model to infer the depth for each pixel directly, in which lots of work followed this line of research [11]. During the training, the dataset includes inputs and correct outputs, which allow the model to learn over time. The algorithm testing data aims to measure the accuracy through the loss function, modification until the error has been sufficiently minimized. Supervised learning problems are separated into two kinds of results: classification and regression.

B. Self-Supervised Methods

Self-supervised methods only require a small number of unlabeled images to train networks to estimate the depth [6, 12, 13]. These methods are gaining depth information automatically by connecting different input methods. Self-supervision methods suffer from the problems of generalization. Models can only work on a minimal set of scenarios similar to the distribution group training.

Eigen et al. [14] Proposed a self-supervising method for estimating depth maps from Siamese network approaches. The technique utilizes the Siamese DispNet [15], ResNet [16], and VGG [17] based network architectures for depth estimation. Moreover, the method predicts multi-scale contrast maps at four scales which were subsequently linked to the output of the previous Decoder layer and the corresponding encoder output using skip connections. The network is trained with RGB depth and ground images at a resolution of 1242 x 375 pixels. The proposed network has the feature of sharing weights decrease their arithmetic operations by halving the network size, leading to a potential consumer model.

Aleotti et al. [18] Proposed a self-supervised system for assessing depth mapping using the residual single end-to-end coordination known as mono-ResMatch. The framework uses the stereo match approach to estimate depth. The feature is assigned an RGB image area and then manufactured to have elements in line with the default correct images. The network also looked at the high-dimensional features at the resolution of the input image to find a multi-scale mirrored depth map aligned with the input image.

The model construction is based on an hourglass structure with skip connections. The final stage consists of a contrast refinement unit that estimates the remaining corrections to the initial variance. The network is trained using Loss of Structural Similarity Index Measure (SSIM), Loss of smoothness contrast with the perceptive term for the edge, and reversal of Huber loss [19]. The model has trained on Cityscapes [20] and KITTI [21] datasets with 640 × 192 random crops.

C. Semi-Supervised Method.

Semi-supervised method has a power to overcome the need to estimate the depth of high-quality images. The semi-supervised approach requires less labelled data and many unlabeled data for training [14, 6]. The disadvantage of semi-supervised methods is that the networks cannot correct their alignment, besides it requires additional field of information such as the focal length of the camera and sensor data.

Cadena et al. [22] proposed GASDA, a semi-supervised method to estimate depth maps using the geometry-aware symmetric domain adaption. This approach targets the generalization of depth estimation methods by training the model on compositional data to estimate depth from natural images. The technique uses the translation of symmetric pattern images and the prediction of monophonic depth. Utilizing the CycleGAN [23], GASDA includes translations of authentic, unreal, unrealistic images and epi-polar geometry of accurate stereo images. The network is trained with two image pattern translations and symmetric depth capabilities to produce depth maps with 192 x 640 pixels.

IV. METHODOLOGY

A. Network Architecture

The algorithm uses a Deep Neural Network to encode-decode depth estimation to estimate the depth for every pixel in an input image. Depth estimation neural network architecture typically consists of an Encoder-Decoder type structure [25]. In existing depth estimation neural networks, the Encoder is often based on an image classification network that may be deep, with numerous layers. Decoders are often design based on the required output and the nature of the dataset and tend to consist of complex building blocks. However, these models are too large and too computationally heavy for real-time inference on devices with resource constraints. The most common types of layers found in in-depth estimation neural networks are convolution layers that perform two-dimensional convolution on an input feature map. Depth estimation DNN shows in Fig. 1. The input here is a three-channel red, green, and blue image of resolution 224 x 224 pixels. The output is a single channel dense depth map of the exact resolution, 244 x 244. Adopt an Encoder-Decoder structure with a low latency design for both.

Encoder network: the blue part shows in Fig. 1 based on the compact and computationally simple Mobile Net Network. It extracts information from the input RGB image from instance features like corners and edges. Popular choices include VGG-16 [24] and ResNet-50 [26] because of their solid, expressive power and high accuracy. Therefore, such networks endure high complexity and latency, making them inappropriate for applications running in real-time on an embedded system. Targeting low latency and efficient state-of-the-art network, Mobile Net [27], applied as an encoder in this experiment. Each layer contains several filters, depending on the number of feature map channels processed by that layer. In a standard convolutional layer, the layer computation involves M times N filters if there are M input channels and N output channels. The height and width dimensions of the filters are typically a small, odd number, such as 3 x 3 filter or 5 x 5 filter. One example is the depth-wise separable layer, which splits the standard convolutional into two.

A depth-wise convolution operates on a channel-by-channel basis. This convolution uses a filter with a height and width of K and a single channel each. A point-wise convolution that scales and sums pixel's value across channels. This convolution uses multi-channel filters, but they have a height and a width of one. The net effect of this depth-wise decomposition is the reduction in parameters due to there being fewer filters to train. The resulting deep neural network is less computationally costly. The depth-wise separable convolution layer was introduced as part of the Mobile Net family of image classification networks. This layer is considered heavily incorporated throughout our own depth estimation neural network.

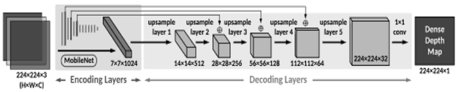


Fig. 1. Proposed network architecture. Dimensions of intermediate feature maps are given [28]

Decoder network: The yellow part shown in Fig. 1 represents the Decoder, which aims to merge and unsampled; it takes the output of the Encoder to form a dense prediction. It consists of five cascading up sample layers. Each of these up samples layers performs a single 5 x 5 convolution. That is followed by nearest-neighbors interpolation. Furthermore, it allows reducing the computation in each up-sample layer by a factor of four.

Skip connection: Encoder networks include numerous layers to gradually minimize the spatial resolution and extract higher-level features from the input. The Encoder and Decoder may be linked even farther through skip connections. It has low-resolution features from the Encoder directly fed into the Decoder, where they are gradually up-sample and merged to perform a high-resolution depth map output. Skip connection permit image details from high-resolution maps in the Encoder to be consolidated into features within the Decoder; this supports the decoding layers reconstruct a more detailed dense output.

B. Loss function

The loss function is used to measure the difference between the ground truth, y, and the expected value of y. It is often referred to as the difference as a prediction error and gives a numerical value to the extent of the network's ability to perform it is mission. When the work forecasts are in pixels, an error calculation and prediction are also in pixels. During the training process, it is updated weights and biases to reduce job loss [28]. Therefore, choosing the appropriate loss function is essential for the network to learn it is task correctly. Some standard loss functions for regression problems are L1 (mean absolute error), L2 (mean square error) Fig. 2.

C. Evaluation's Metrics and Model Evaluation Criteria.∑

The following metrics measure the depth error. Assume p is the prediction of pixel depth in the image, g is the base fact, and N is the total number of depth pixels.

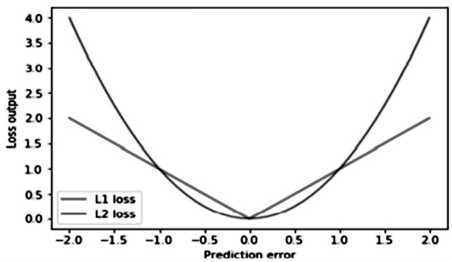


Fig. 2. The L1(mean absolute error) and L2 (mean square error) losses as a function of the prediction error, y − yˆ.

TABLE II. THE METRICS EQUATIONS FOR EACH METRICS METHODS

Metrics	Equation	
Relative Error	RE= $\frac{1}{N}\sum_i\sum_j\frac{\lvert g_{ij}-P_{ij}\rvert}{g_{ij}}$	(1)
Squared Relative Error	SRE= $\frac{1}{N}\sum_i\sum_j\frac{\lvert g_{ij}-P_{ij}\rvert^2}{g_{ij}}$	(2)
Root Mean Squared Error	RMSE= $\sqrt{\frac{1}{N}\sum_i\sum_j(g_{ij}-P_{ij})^2}$	(3)
Logarithmic Root Mean Squared Error (logRMSE)	logRMSE= $\sqrt{\frac{1}{N}\sum_i\sum_j(\log(g_{ij})-\log(p_{ij}))^2}$	(4)

D. KITTI Dataset

In particular, several important datasets estimate the depth of the images because they provide images and maps of the corresponding depth from different measurable points. The deep learning method of depth estimation requires more and more training and assessment datasets that contain heterogeneous data such as images, videos, ranges, etc. Therefore, it is essential to define a suitable and high-resolution dataset for real-time detect objects and estimate the depth of the applications of roads and railways.

For training and evaluating depth-predicting networks, the KITTI dataset [15] is a popular choice. The dataset consists of about 93,000 images of outdoor scenes, divided into five categories: City, residential, road, campus, and person. Data recorded with a high-resolution stereo color camera and a Velodyne 3D laser scanner (LiDAR), which captured the scene's depth. RGB's resolution varies slightly depending on the calibration parameters but is around 1242 x 375.

The available depth maps are four times denser than the initial LiDAR scans. That is, they have values equivalent to four times the number of pixels. Primary LiDAR scans are few, which means it does not have values for all pixels. However, Ma et al. [28] provided more intense depth maps, which derived an extensive dataset of deep annotated RGB images from scattered scan data. They stimulated their work by showing that convolutional neural networks typically perform low when trained on sparse data. Fig. 3&4 show an example of an input image and the corresponding ground truth depth map from the KITTI depth dataset [28]. Only valid values are available for the lower part of the image for all depth maps, and invalid pixels (pixels without a depth value) have zero value.

The image shows in Fig. 3; the image is processed to display depth in meters; the color of each pixel represents the depth value. Yellow represents large values and purple for small values. Invalid pixels that lack depth information have a value of zero and it characterizes by black. Note that the depth map shown in Fig. 4 does not contain any valid values for the upper part.

Fig. 3. Example of a synced and rectified RGB Image from the KITTI dataset [19].

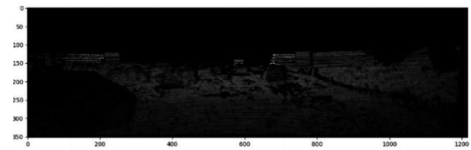


Fig. 4. Corresponding depth map [31]

V. Experimental Results and Analysis

In this section, evaluate the experiment results to demonstrate the impact of parameters on monocular depth estimation.

A. Training

According to previous research [29, 30, 31], the network implemented in TensorFlow [32] contains 31 million trainable parameters. It takes 30 hours to train with a single Titan X GPU on a dataset of 29,000 images for 50 epochs. The inference is fast and takes less than 35ms, or more than 28 frames per second, for a 512 x 256 image, including transfer times to and from the GPU.

A baseline model is used to train particularly 200 selected images; the training of this model is done on a local machine equipped with the Nvidia RTX2060 GPU processor. The training's hyperparameters for the baseline model used as stated in [29]. A series of trials conducted later with different hyperparameters evaluated the model through metrics that their values change by changing the hyperparameters. Table III explains the set of training trails with each training parameter and the time consumed.

B. Model Evaluation

Six experiments were conducted to check the effects of hyperparameters on the model output. The default hyperparameter [30] set in the first experiment (E1), then the selected parameters changed respectively, as shown in Table III.

TABLE III. The Hyperparameters Selected for Experiments

Experiments	Epochs	Learning Rate	Loss Weight	Training Time, *h*
E1	50	0.0001	1	8
E2	30	0.00005	0.55	4
E3	45	0.00005	0.55	7
E4	30	0.00001	0.55	4
E5	30	0.00001	0.01	4
E6	30	0.000005	0.01	4

The results of experiments show in Table IV explain the significant impact of learning rate and loss weight values on the training process for the same model. Hence, based on Table IV, experiment number four refers to the best hyperparameters values.

TABLE IV. Evaluation Metrics Results for Each Experiment

Experiments	Abs_rel	Sq_rel	RMS	Log_rms
E1	0.2662	3.918	9.11	0.371
E2	0.2594	3.466	8.82	0.350
E3	0.2561	3.701	8.97	0.358
E4	0.2449	2.616	8.19	0.341
E5	0.2568	2.944	8.83	0.369
E6	0.2792	3.733	9.05	0.370

VI. Conclusion

This paper presented an overview of the previously implemented model to estimate the depth from monocular images and a short explanation of the main methods used to develop the depth estimation model. The network architecture of the depth estimation model is mainly built up of Decoder and Encoder layers. Besides that, an experiment has been run to achieve the objective of this paper, a baseline model built based on a state-of-art depth estimation model using the default hyperparameters, followed by five experiments to tune the learning rate and loss weight. The results show that the high training accuracy achieved at learning rate equal 1e-5 and loss weight equal 0.55 with less time consumed for training.

References

[1] C. Godard, O. M. Aodha and G. J. Brostow, "Unsupervised Monocular Depth Estimation with Left-Right Consistency," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6602-6611, doi: 10.1109/CVPR.2017.699.

[2] S. Saleh and S. A. Khwandah, 'Traffic Signs Recognition and Distance Estimation using a Monocular Camera', 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 407–418, 2019.

[3] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari and N. Navab, "Deeper Depth Prediction with Fully Convolutional Residual Networks," 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 239-248, doi: 10.1109/3DV.2016.32.

[5] M. A. Islam, N. Bruce and Y. Wang, "Dense Image Labeling Using Deep Convolutional Neural Networks," 2016 13th Conference on Computer and Robot Vision (CRV), 2016, pp. 16-23, doi: 10.1109/CRV.2016.31.

[6] L. Chen, A. Hermans, G. Papandreou, F. Schroff, P. Wang and H. Adam, "MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction Features," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4013-4022, doi: 10.1109/CVPR.2018.00422.

[7] R. Ji et al., "Semi-Supervised Adversarial Monocular Depth Estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 10, pp. 2410-2422, 1 Oct. 2020, doi: 10.1109/TPAMI.2019.2936024.

[8] C. Godard, O. M. Aodha, M. Firman and G. Brostow, "Digging Into Self-Supervised Monocular Depth Estimation," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3827-3837, doi: 10.1109/ICCV.2019.00393.

[9] S. Saleh, S. Manoharan, and W. Hardt, “Real-time 3D Perception of Scene with Monocular Camera,” Embed. Selforganising Syst., vol. 7, no. 2, pp. 4–7, Sep. 2020.

[10] Y. Cao, Z. Wu and C. Shen, "Estimating Depth From Monocular Images as Classification Using Deep Fully Convolutional Residual Networks," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 11, pp. 3174-3182, Nov. 2018, doi: 10.1109/TCSVT.2017.2740321..

[11] B. Ummenhofer et al., "DeMoN: Depth and Motion Network for Learning Monocular Stereo," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5622-5631, doi: 10.1109/CVPR.2017.596.

[12] S. Zhao, H. Fu, M. Gong and D. Tao, "Geometry-Aware Symmetric Domain Adaptation for Monocular Depth Estimation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9780-9790, doi: 10.1109/CVPR.2019.01002.

[13] W. Yin, Y. Liu and C. Shen, "Virtual Normal: Enforcing Geometric Constraints for Accurate and Robust Depth Prediction," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2021.3097396.

[14] D. Eigen and R. Fergus, "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-scale Convolutional Architecture," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2650-2658, doi: 10.1109/ICCV.2015.304.

[15] N. Mayer et al., "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4040-4048, doi: 10.1109/CVPR.2016.438.

[16] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[17] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

[18] F. Tosi, F. Aleotti, M. Poggi and S. Mattoccia, "Learning Monocular Depth Estimation Infusing Traditional Stereo Knowledge," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9791-9801, doi: 10.1109/CVPR.2019.01003.

[19] T. Qin, P. Li and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," in *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004-1020, Aug. 2018, doi: 10.1109/TRO.2018.2853729.

[20] M. Cordts et al., "The Cityscapes Dataset for Semantic Urban Scene Understanding," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3213-3223, doi: 10.1109/CVPR.2016.350.

[21] S. Zhao, H. Fu, M. Gong and D. Tao, "Geometry-Aware Symmetric Domain Adaptation for Monocular Depth Estimation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9780-9790, doi: 10.1109/CVPR.2019.01002.

[22] C. Cadena et al., "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," in IEEE Transactions on Robotics, vol. 32, no. 6, pp. 1309-1332, Dec. 2016, doi: 10.1109/TRO.2016.2624754.

[23] J. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

[24] K. Simonyan and A. Zisserman, "Intense convolutional networks for large-scale image recognition," International Conference on Learning Representations (ICLR), 2015.

[25] S. Saleh, S. Manoharan, J. Nine and W. Hardt, "Towards Robust Perception Depth Information For Collision Avoidance," 2020 IEEE Congreso Bienal de Argentina (ARGENCON), 2020, pp. 1-4, doi: 10.1109/ARGENCON49523.2020.9505483.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770– 778.

[27] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[28] D. Work, F. Ma, T. J. Yang, S. Karaman, and V. Sze, "FastDepth: Fast monocular depth estimation on embedded systems," arXiv, p. 40, 2019.

[29] F. Ma, G. V. Cavalheiro and S. Karaman, "Self-Supervised Sparse-to-Dense: Self-Supervised Depth Completion from LiDAR and Monocular Camera," 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 3288-3295, doi: 10.1109/ICRA.2019.8793637..

[30] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 6602–6611, 2017.

[31] S. Shadi, S. Hadi, M. A. Nazari, and W. Hardt, ‘Outdoor Navigation for Visually Impaired based on Deep Learning’, 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 397–406, 2019.

[32] Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep learning*. MIT press.

Deep-Learning-Based Insulator Detector for Edge Computing Platforms

Batbayar Battseren
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
batbayar.battseren@informatik.tu-chemnitz.de

Mohamed Salim Harras
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
mohamed-salim.harras@informatik.tu-chemnitz.de

Soaibuzzman
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
soaibuzzman@s2019.tu-chemnitz.de

***Abstract*—In the past years, the object detection applications have witnessed a rapid increase in usage of deep-learning (DL) based solutions, due to their accurate object detection, and robustness to illumination, scale, clutter, rotation changes, etc. Therefore, DL-based approaches started to be used in real-time applications. In an autonomous aerial inspection system, the robust detection and time requirement are critical aspects for the real-time perception and decision making. However, most of the DL models are not suitable for the edge computing platforms, due to their heavy sizes and poor reliability. This paper presents the experimental results obtained from a YOLOv4-Tiny model with a CSPDarknet-53 backbone on different single board computers. The study demonstrates that the performance of the adopted approach is highly dependent on the target platform; and the real-time object detection is reachable in specific cases.**

Keywords—deep learning, object detection, power line inspection, insulator, embedded device, edge computing, autonomous system, unmanned aerial vehicle

I. Introduction

Unmanned aerial vehicles (UAV) have been used in the past years for different kinds of critical infrastructure inspection, due to the benefits that they offer. The UAV-based inspection provides a number of advancements in the high voltage power line inspection process compared to the conventional methods. For instance, no human life is at risk, no voltage cut-offs, high-quality data collection, high accessibility, less time consumption, and lower cost. However, there are still some weaknesses in the UAV-based inspection which are mainly caused by external factors like the weather condition, the environmental difficulty, or by the pilot experience, response time, or fault.

The power line inspection is a highly automatable sequential process, and most of the object inspections can be performed using vision cameras. The Autonomous Power Line Inspection (APOLI) project addresses the abovementioned problem by offering an autonomous solution for these processes [1]. The advantages are that the autonomous system can perform repeated processes without any problem, have a better control and response time in strong environmental conditions, and provide a better quality inspection.

The unmanned aerial system (UAS) that was developed consists of three main components, which are the copter, the camera setup, and the companion computer. Technically, the copter could be any type of multi-copter that has enough payload to carry the autonomous system's additional components. The flight controller should be open source or controllable by external commands. The camera setup is equipped with navigation and inspection cameras which are mounted on a 3-axis gimbal. The dual-camera configuration delivers simultaneously a wide-angle navigation view and a narrow-angle high-resolution inspection view. The companion computer is directly mounted on the copter in order to run the image processing and autonomous decision-making algorithms in real-time.

In order to develop this type of autonomous system, we proposed an Adaptive Research Multicopter Platform (AREIOM) destined for being applied within the APOLI project [2]. Fig. 1 illustrates the AREIOM software and hardware architecture, where the software components are mapped on a dedicated hardware component. There are several software components developed to execute the inspection mission automatically, namely Expert System (EXS), MAVLink Abstraction Layer (MAL), Camera Gimbal Control (CGC), Acquisition (ACQ), Navigation Image Processing (NIP), and Inspection Mission Recorder (IMR).

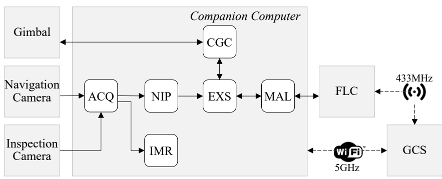


Fig. 1. AREIOM software and hardware architecture

The EXS plays a leading role in running the inspection mission and making real-time decisions [3], while the MAL sends and receives messages from the Flight Controller (FLC) [4]. Together, they work as a flight subsystem. At the same time, the ACQ, the NIP, the IMR, and the CGC are acting as a vision subsystem. The ACQ acquires the live video feed from the cameras and provides it to other processes like the NIP and the IMR. The CGC points the cameras into the desired direction with the help of the NIP output. The NIP detects the inspection object, and the IMR records high-resolution inspection data.

In this paper, we discuss the navigation image processing part, specifically the deep-learning (DL) based insulator detection, including the model training and the experiments performed on different embedded devices.

II. RELATED WORK

Over the past decade, the DL approaches for object detection have seen a consistent increase. This increase was met with a high demand for inspection routines using UAVs equipped with cameras. With the help of edge computing devices and high-performing cameras, inspection routines based on DL became a topic of active research lately. Throughout the DL training and learning of image features, structures are detected accurately and at faster speeds. UAV-based autonomous inspection system requires a continuous stable detection of the insulator string in order to control the flight and inspect the insulators.

A. Feature-Based Insulator Detection

We developed a traditional image processing algorithm for insulator detection, which detects the target object without any machine learning method and parallelism [5]. This approach detects the insulator string based on its unique features, which are symmetrical shape and repeated color pattern. Initial localization of the insulator takes place based on symmetry analysis. The aforementioned analysis extracts image features (pixel intensities, edges) and analyses the edges for symmetry in the Hough space. In the study, we tested the approach on lab images with a solid white background which led to a robust detection of insulators. The same approach was developed further for real-life images in [6]. To achieve the same result in a real-life scenario without using the DL, a color feature-based signal representation analysis method was introduced. In addition to the previous method, this approach detects the repeated color pattern of the insulators on the symmetry axes, which completes the region of interest (ROI) of the insulator. Moreover, an insulator burn-mark damage detection method in [7] was improved and used in the real-life scenario in this study.

The measured average detection frame rate was 15 fps on the machine with an i5 CPU and 8GB RAM configuration. Though this method was computationally expensive and very slow in terms of performance, it has been tested on the Odroid XU4 single board computer (SBC), and the average detection rate was only 5.48 fps. It also led to many false detections since the color feature could be affected by any external factors like illuminance, view angle, reflection, camera configuration, etc.

Jabid et al. [8] defined an approach based on spatial features. The method makes use of morphological operations and defines a spatial model for the detected object. In their method, the concept of a sliding window based local directional pattern (LDP) feature is extracted and support vector machines are used for the classification of the extracted windows.

B. Deep-Learning-Based Insulator Detection

In the past year our team have already conducted several DL based solutions in related studies and projects [9, 10]. In these studies, DL were used in collision warning, outdoor navigation, and depth estimation application. In [11], DL-based single insulator detector approach presented. In the Master Thesis report published at Chemnitz University of Technology [12], a student proposed a DL-based approach for burn-mark detection. A YOLOv5 network has been chosen, and the learning stage is performed with both passive and active approaches. The active learning approach allowed to have a lightweight model with similar or higher accuracy.

Naeem et al. [13] presented a DL-based autonomous vision system to detect faults on insulators based on a YOLOv4-tiny network architecture [14]. To evaluate the performance of their model, the YOLOv4-tiny results were compared to YOLOv3-tiny network architecture on different edge-computing devices such as the Raspberry Pi 4, Nvidia Jetson Nano, Nvidia Jetson TX2, and Nvidia Jetson AGX Xavier. They concluded, using YOLOv4-tiny, real-time detection can be realized on-board of Nvidia Jetson AGX Xavier without a trade-off of accuracy.

Adou et al. [15] proposed a method for insulator bunch-drop detection based on deep learning. The authors opted for the YOLOv3 model to train their network. In fact, two labels were defined for the insulator and the bunch-drop. The model classifies the detected objects and allows to localize the insulator. Logistic regression is the tool used for performing classes probabilities and labels predictions within the YOLO model. The proposed method was able to perform well on a desktop, processing frames at 45 frames per second with a mAP of 83,52%.

In [16], the authors defined a cascading approach composed of the detection network and the classification network. The first network allows to detect the insulators based on RPN, which are then fed to the classification network for fault detection. This allowed for the reduction of the computational cost, while the accuracy and the robustness of the model improved considerably.

A different study targeted the detection of the insulator using the YOLOv2 model [17]. The proposed method used data augmentation to increase the size of their dataset and avoid over-fitting. Hence, they recorded a mAP of 88% with an average prediction time of 0,04 seconds, allowing them to run at real-time.

A deep CNN cascading architecture is introduced in [18] based on the concatenation of VGG and ResNet in a Region-Based Convolutional Neural Network (R-CNN) for the localization of defects on insulators. The cascading architecture allows for two-stage object detection. The localization of the insulator is based on a region-based proposal (RPN) approach and is then followed by the localization of defects in the proposed regions. Another research [19] based the localization of the insulator on a Faster R-CNN, an improved version of R-CNN that cuts the running time of the RPN [20].

In a benchmark published by Nvidia [21], several models with their image resolutions have been tested for inference on most edge computing devices produced by Nvidia, and the ResNet network along with YOLOv4-tiny and SSD (Single-Shot Detection) networks perform best on embedded devices.

In contrast to two-stage detection methods, Lui et al. [22] propose a single-shot detector (SSD) based on the VGG16 model for feature extraction. The latter bases its prediction on 21 class scores for every detected instance. By excluding the region proposal network (RPN), the network is able to meet the real-time requirements while allowing for fair detection accuracy.

III. Deep Learning Model Training

Fig. 2 illustrates the main steps of the DL-based insulator detector implementation. The data preparation and model training points are discussed in this section. The model deployment is discussed in detail in the next chapter.

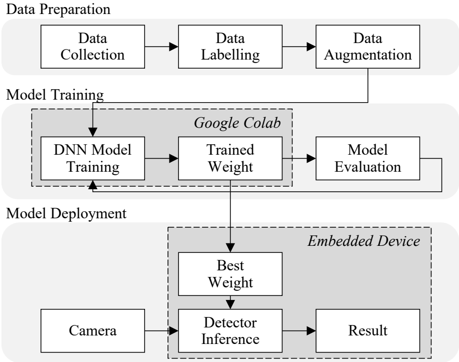


Fig. 2. Workflow of the DL-based insulator detector implementation

A. Dataset Preparation

In the APOLI project, the main inspection object is the high voltage power line insulator, which is an essential part in the power pole. It holds the wire without making any electrical connection with the power pole. The insulators are generally linked together to form what is called an insulator string, and the number of the insulator is defined by the voltage level. Though there are different types of insulators available on the market, the specific insulator targeted in this study is the PS-70 glass insulator, which is commonly used in many countries (Fig. 3).

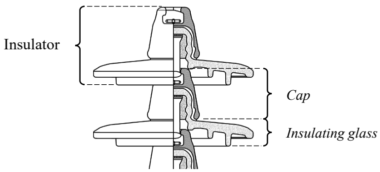


Fig. 3. High Voltage Power Line Insulator

From the computer vision perspective, insulators have their unique features. As they are rotating objects, they have symmetric representation on the image. Furthermore, as they are partially made of glass, they are reflective, and their color changes depending on the environment and the view angle.

1) Data collection

According to empirical bounds for training datasets on computer vision, a rule of thumb is 1,000 images per class [23], though this number can be reduced if pre-trained weights are used. The image resolution is 640×640 and the dataset is collected from real high voltage power line inspection videos, which are recorded under the APOLI project from 2016 to 2019. The whole dataset is categorized into four distinct categories based on their quality and captured environment: good, average, bad, and indoor. The good dataset contains a huge diversification on images with different angles and backgrounds in good quality, whereas the average and bad datasets hold less quality and blurry images. The indoor images are captured in the lab in different lighting conditions. This variation of the dataset helps to make the model more robust, hence, to deal with the over-fitting problem.

After that, images are selected randomly from the training set based on a constant selection ratio. This ratio is selected heuristically: 35% of the image from the good category, 25% of the image from the average category, 20% of the image from the bad category, and 20% of the image from the indoor category.

2) Data augmentation

In order to provide more robustness to the network, data augmentation techniques are applied to the selected images. Detecting the insulator at different angles is a significant challenge when these kinds of images are not included in the dataset. This challenge is overcome by rotating and flipping the training images at a certain angle. Moreover, a salt and pepper filter was applied to make the model robust against the high-frequency noise. The following list depicts the three chosen augmentation techniques:

- Rotation: Both 90° clockwise and anti-clockwise rotations are applied to create scenes with horizontal insulator strings. The dimension is preserved after rotation.
- Flip: Only horizontal flip is applied. This is mainly to create scenes with inversed insulators.
- Noise: A high pass filter is used. The purpose of this filter is to create scenes where the UAV's camera feed is blurred by sudden UAV movements. A 2% salt-and-pepper filter is applied to each image to distort the high-frequency features.

A canvas of the original image and the augmented images are illustrated in Fig. 4.

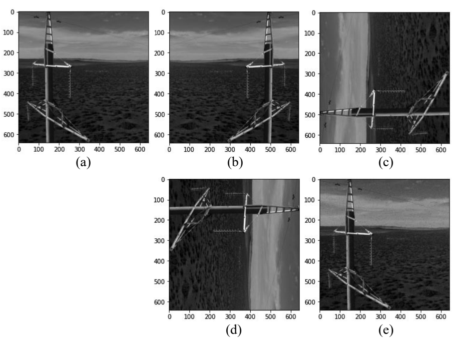


Fig. 4. Illustration of image samples with augmentation, (a) original image, (b) horizontal flipped image, (c) 90° anti-clockwise and (d) 90°clockwise rotation, and (e) 2% salt & pepper noise is added in the original image.

B. Model Training

Being computationally extensive, region proposals-based two-step object detection algorithms are not feasible for resourced constrained embedded systems. By compromising some accuracy, it is possible to gain a significantly higher inference speed with a regression-based one-stage approach. There are a few prominent one-stage object detection algorithms are available out there such as SSD and YOLO. To meet the edge computing requirements, a trade-off in accuracy is unavoidable.

1) Model Selection

Nvidia Jetson Nano is considered as the main target platform for the inference phase in our application. An inference benchmark for DL-based object detection methods on Jetson Nano is developed by the Nvidia developers' community [24]. In this benchmark, SSD with a mobilenet-v2 backbone performed the inference at a faster speed compared to Tiny YOLOv3. Though the inference speed is a bit higher for SSD, the Tiny YOLOv3 showed a much better performance in terms of accuracy with a mean average precision (mAP) of 0.70 at an intersection over union (IoU) threshold of 0.5 on the MS COCO dataset [25]. The further success of the YOLO family, YOLOv4-tiny [26], allowed for an increase in the usability of object detection applications in edge computing systems.

The size and required computational power for YOLOv4-tiny are much smaller compared to YOLOv4. It uses the CSPDarknet53-tiny network as the backbone which uses the CSPBlock module in cross SPN instead of the ResBlock module in the residual network of CSPDarknet53 (Fig. 5). Hence the numbers of convolutional layers are compressed. However, considerable precision is still maintained. To make the computation procedure more efficient, the LeakyReLU activation function is used in the YOLOv4-tiny instead of the Mish activation function. To increase the object detection speed, rather than using YOLOv4's SPP and PAN, it uses two different scales feature maps that are 13×13 and 26×26 to predict the detection results.

The adopted methodology for the study is based on a YOLOv4-tiny network, a regression-based one-step object detection algorithm, and the CSPDarknet53-tiny model backbone.

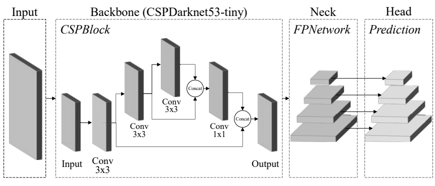


Fig. 5. Architecture of proposed YOLOv4-tiny model with CSPDarknet53-tiny

2) Training Configuration

To build the CSPDarknet53-tiny environment some dependencies need to be fulfilled such as – OpenCV, Cuda Toolkit, cuDNN, and GPU architecture. To enhance this process, Google Colab with Tesla K80 GPU is used. The pre-trained weights of YOLOv4-tiny are used for accelerating the training process and getting better accuracy. A few other configurations have been done to train the model. This optimal configuration has been made based on the official GitHub repository of YOLOv4. Hence, the model architecture is adjusted depending on the number of defined classes in the dataset. The configurations are:

- Batch = 64
- Subdivision = 16
- Max batches = classes×2000, but not less than the number of training images and not less than 6000.
- Steps = 80% and 90% of the max batches i.e., steps=4800; 5400.
- Filters = (classes + 5) × 3 i.e., 18 for our one class.
- Width = 416 & Height = 416.
- Learning rate = 0.00261

The model is trained for 6000 epochs with this configuration. There are 9 weight files created once the training process is completed. One for every 1000 epochs, hence 6 in total. The other 3 weight files are – the best, final, and the last. We used the best weights file for our inference which provides the highest accuracy.

C. Model Evaluation

For the model evaluation, confusion matrix-based metrics are used. There are four main terms in the confusion matrix, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which are used for calculating the accuracy, precision, recall, and F1-score. The accuracy presents the number of correct classification over the total number of data (1). The precision represents what percentage of the positive prediction were true (2). The recall calculates what percentage of the actual positive were predicted correctly (3). The F1-score combines the precision and recall into a single metric (4). In addition, IoU score is used as an evaluation metric (5) measures the detection accuracy based on the ground truth and the prediction.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (5)$$

Fig. 6 illustrates the calculated loss and mean average precision (mAP) of the proposed model during the training period. The combination of the precision and recall curve is called the precision-recall curve (PRC). The area under the PRC expresses the mAP. The higher the value of mAP, the better the accuracy of the model. The red line represents the mAP over the training iteration, while the blue line represents the loss. The mAP was 93,7 % after 6000 iterations. The binary cross-entropy is used for the loss function and the average loss was 0.0702.

In order to evaluate how the model performs, three different unseen datasets were prepared. The first dataset is collected from the inspection videos, which are recorded by Yuneec E90 high-resolution camera, while the second dataset is taken with FLIR Blackfly S camera. The third dataset is gathered from an indoor setup with a FLIR camera. In order words, the first two datasets are collected from the same environment with different cameras, while the second and third datasets are collected from a different environment with the same camera. Fig. 7 presents the confusion matrix of the trained model under different IoU thresholds on the first dataset.

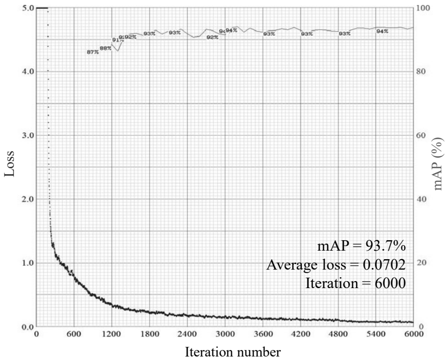


Fig. 6. Loss and mAP graph, red line mAP curve, blue line Loss curve

The trained model has been evaluated on the prepared datasets under different IoU thresholds. For the evaluation 0.5, 0.75, 0.85, and 0.9 IoU thresholds were used to measure the defined metrics. The model evaluation result is summarized in 0 By analyzing the data, it is possible to conclude that the current model performs excellently on 0.5 IoU threshold with mAP of 78.6% – 98.9% along with all datasets. However, as the IoU threshold increases, the model accuracy drops significantly. It is possible to achieve a higher mAP on a larger IoU threshold by increasing the dataset and longer training. But that will affect the inference time.

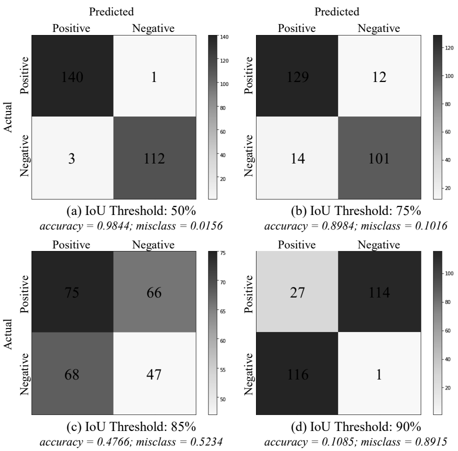


Fig. 7. Confusion matrix results for Dataset 1, (a) IoU threshold at 50%, (b) IoU threshold at 75%, (c) IoU threshold at 85%, and (d) IoU threshold at 90%.

TABLE I. MODEL EVALUATION RESULT

	Dataset 1				Dataset 2				Dataset 3			
IoU thr.	0.50	0.75	0.85	0.90	0.50	0.75	0.85	0.90	0.50	0.75	0.85	0.90
Accuracy	0.98	0.89	0.47	0.11	0.90	0.36	0.16	0.14	0.84	0.51	0.28	0.17
Precision	0.98	0.90	0.52	0.19	0.85	0.43	0.17	0.08	0.82	0.44	0.18	0.06
Recall	0.99	0.91	0.53	0.19	0.98	0.49	0.20	0.09	0.82	0.44	0.18	0.06
F1-score	0.98	0.66	0.53	0.19	0.91	0.46	0.19	0.09	0.82	0.44	0.18	0.06
IoU	0.83	0.77	0.46	0.17	0.64	0.36	0.15	0.07	0.61	0.37	0.16	0.05
mAP (%)	98.9	89.7	32.4	5.01	97.0	24.9	3.96	0.85	78.6	27.0	4.80	0.62

IV. EXPERIMENTATION

A. Hardware setup

The experiment has been done on the copter which was developed according to the AREIOM architecture. Besides the copter itself, the main hardware components are the companion computer and the camera setup. To make the UAS lightweight, compact, portable, and low-power consuming, the vision subsystem is realized as an embedded system. For this reason, the FLIR industrial cameras have been selected for navigation and inspection purposes due to their low weight, compact size, and reasonable performance. The navigation camera is a 3.2 Megapixel global shutter FLIR Blackfly S camera with a Sony IMX252 sensor, which is equipped with Tamron 3-14 mm focal length zoom lens. The camera maximum resolution is 2048 x 1536 pixels, the maximum frame rate is 55 fps, and the lens field of view (horizontal × vertical) is from 105.4°×77.6° to 33.0°×24.8°. The camera uses a USB 3.0 interface, and the maximum throughput is limit to 380MBps.

The experiment has been carried out on two different SBCs, which are Nvidia Jetson Nano and Nvidia Jetson Xavier NX. Both of them have the necessary CUDA cores, high-speed interface, and the same 64-bit architecture. The Jetson Nano has a quad-core ARM Cortex-A57 CPU, 128 CUDA core Nvidia Maxwell GPU, 4 GB 64-bit LPDDR4 memory, and four USB 3.0 interface. Due to the GPU performance, USB3.0 interface, and small form factor, this device is the best low-cost option. On the other hand, the Jetson Xavier has a 6-core Nvidia Carmel ARM CPU, 384 CUDA core Nvidia GPU, 8 GB 128-bit LPDDR4, and four USB 3.1 interfaces, which makes it a better performing alternative with a higher budget.

B. Experiment on the embedded platform

In this section, the trained model's inference time is mainly discussed. The insulator detection model has been tested on the two target platform in different power modes. On the Jetson Nano 5W and 10W power modes, and on the Jetson Xavier NX 10W, 15W, and 20W power modes were used for the experiment (see TABLE II.). The image processing can be divided into image acquisition and image processing steps.

1) Image acquisition

In order to acquire frames from the FLIR industrial cameras, the Spinnaker Software Development Kit is used. Spinnaker is a Generic Interface Camera (GenICam) based Application Programming Interface, which supports FLIR's USB vision cameras and provides necessary functionalities.

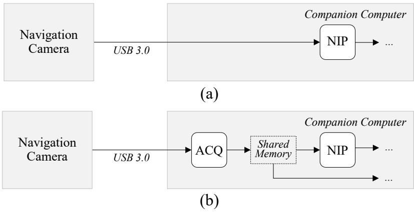


Fig. 8. Image acquisition solutions, (a) Integrated acquisition, (b) Modular acquisition

Fig. 8 illustrates the two image acquisition methods, which are proposed for the system. Each of them has its advantages and disadvantages.

The first option (Fig. 8(a)), the integrated acquisition is a straightforward solution that does not require any additional component and is relatively simple to implement. The NIP acquires images from the camera and directly applies the object detection model to it as a single process. However, this integrated implementation of the image acquisition and image processing would not allow other components to access the same image feed simultaneously.

On the other hand, the second option (Fig. 8(b)), the modular acquisition has relatively complex implementation with separate components and provides a number of possibilities to the system compared to the other alternative. The main difference is the data flow of the image data. The first component, Acquisition (ACQ) configures the camera, reads the video frames with the help of the Spinnaker library, and writes them to the Shared Memory as a byte array. The shared memory is used for sharing the data for other processes, due to some advantages like no data are explicitly moved, high-speed data access, and no overhead. The blackboard architecture is used to implement this data-sharing mechanism [27]. In a nutshell, the blackboard architecture has three main components, namely the Blackboard, which is a common memory space where the data are stored, and the Knowledge Sources, which are the processes that read and write data into the memory. Lastly, the Control component selects the knowledge sources to interact with the blackboard. There can be any number of knowledge sources. According to the AREIOM architecture, every software component is a knowledge source. The second component NIP reads the frames from the blackboard and applies the object detection model to it. This makes the second solution more modular by separating the image acquisition part from the image consumer components. In addition, there can be multiple detection modules and other components which are reading from the shared memory at the same time.

The comparison of the both acquisition approaches results is presented in the TABLE II. In the integrated acquisition approach, the average image acquisition times were 2.042 ms on Jetson Nano and 1.811 ms on Jetson Xavier NX. In contrast, in modular acquisition implementation with shared memory, the average acquisition times were 0.056 ms on Jetson Nano and 0.077 ms on Jetson Xavier NX. The shared memory based implementation decreased the acquisition time by 1.7 ms to 1.9 ms. However, there is no big advancements depending on the power modes.

TABLE II. EMBEDDED PLATFORM TEST RESULT

Device	Jetson Nano (128 CUDA cores)			Jetson Xavier NX (384 CUDA cores)			
Integrated Acquisition							
Power mode	5W	10W	Avg.	10W	15W	20W	Avg.
Acquisition [ms]	2.246	1.839	2.042	2.057	1.723	1.655	1.811
Detection [ms]	111.1	75.67	93.39	24.09	19.739	18.42	20.75
FPS	9.0	**13.21**	11.10	41.50	50.66	**54.28**	48.81
Modular Acquisition							
Power mode	5W	10W	Avg.	10W	15W	20W	Avg.
Acquisition [ms]	0.062	0.050	0.056	0.083	0.079	0.070	0.077
Detection [ms]	112.8	73.48	93.17	24.26	19.45	18.82	20.84
FPS	8.859	**13.60**	11.23	41.20	51.39	**53.13**	48.57

2) Image Processing

There are two possible approaches to predict the target object with our trained YOLOv4 model: the OpenCV DNN module and the Darknet module. Since OpenCV DNN module is optimized for Intel processors, the inference time of the insulator detector was 130.0 ms per frame on the ARM processor. In contrast, the inference time of the Darknet module was 73.48 ms per frame. Accordingly, the Darknet module is selected for further test. TABLE II. presents the measured inference time of the same insulator detector model on different embedded devices. The same trained model performed differently on different circumstances. For example, the detection time was 112.8 ms on Jetson Nano in 5W mode, while it was 20.84 ms on Jetson Xavier NX in 20W mode.

In addition, under the APOLI project, the traditional image processing algorithm has been developed for insulator detection, which is compared with the current DL-based approach on TABLE III. The traditional image processing approach was tested on the Odroid XU4 and Odroid H2 devices, particularly on a CPU. This algorithm is developed without parallelism techniques for running on the CPU cores. Likewise, the DL model is designed to run on the GPUs, and the experiment has been done only on the devices equipped with GPU. The Google Colab experiment result is also included in TABLE III. , which is performed on the Tesla K80 graphic card.

TABLE III. COMPARISON OF THE INSULATOR DETECTION ALGORITHM ON EMBEDDED PLATFORM

Detection Method	Traditional Image Processing		Deep-Learning based Model YOLOv4 tiny		
Device	Odroid XU4	Odroid H2	Jetson Nano	Jetson Xavier	Google Colab
Processing unit	CPU	CPU	GPU	GPU	GPU
CUDA cores	-	-	128	384	4992
Image source	Camera	Camera	Camera	Camera	File
Image resolution	1024x768	1024x768	1024x768	1024x768	416x416
Acquisition [ms]	7.140	1.140	0.050	0.070	0.106
Detection [ms]	175.330	65.030	73.489	18.820	14.743
FPS	5.480	15.110	**13.607**	**53.134**	67.344

The total frame rate of the application is defined by the summation of the acquisition and detection time. Though the detection time takes the biggest portion, as it decreases from 111.1 ms to 18.42 ms (see TABLE II.), the acquisition time ratio increases from 1.98% to 8.24% of the total time.

However, the shared memory-based implementation has shown significant advancement by decreasing the acquisition time by 96.33% on average.

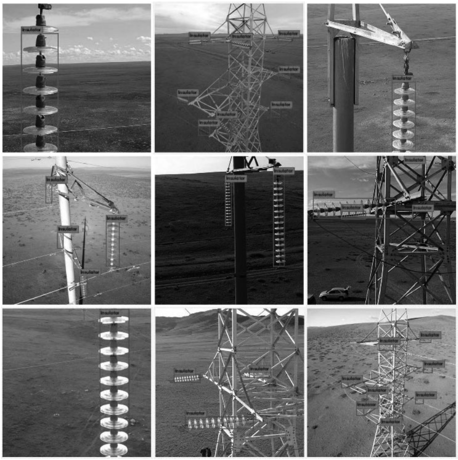


Fig. 9. Detection result of high voltage power line insulators

Fig. 9 shows the detection result of the YOLOv4-tiny insulator detector in different conditions.

Conclusion

In this study, the DL-based insulator detector solution developed, and the performance measured and compared on different edge computing devices. In the training phase, we have carefully collected the training dataset to reach higher accuracy and robustness with fewer training data. Furthermore, YOLOv4-tiny architecture is selected for the implementing the real-time insulator detector for the autonomous system.

The lightweight YOLOv4-tiny detector showed promising results on the selected platforms even without any optimization technique. The proposed methodology was able to reach a maximum frame rate of 13.607 fps on a Jetson Nano board. Meanwhile, it recorded a frame rate of 53.134 fps on the Jetson Xavier NX SBC. In order to reach the aforementioned frame rate, shared memory based data sharing method proposed in the image acquisition stage, and it outperformed the conventional method. Based on the obtained data, the hardware configuration, such as the number of CUDA core and the enabled power mode, have a significant impact on the model's performance.

On the other hand, the training approach, the optimization method, and the dataset are also relevant factors that influence the final model accuracy and the detection time. For this reason, our ongoing research will be targeting the usage of advanced training methods, and the optimization of the current model to decrease the inference time and increase the accuracy. In addition, the ongoing research will compare different DNN architectures to find the optimal model respecting the speed/accuracy tradeoff for the autonomous aerial inspection system.

References

[1] U. Tudevdagva, B. Battseren, W. Hardt, S. Blokzyl and M. Lippmann, "UAV Based Fully Automated Inspection System for High Voltage Transmission Lines," 12th International Forum on Strategic Technology (IFOST 2017), Ulsan, Korea, May 2017

[2] R. Harradi, B. Battseren, A. Heller and W. Hardt, "AREIOM: Adaptive Research Multicopter Platform," 14th International Forum on Strategic Technology (IFOST 2019), Tomsk, Russia, pp. 219-223, Oct 2019

[3] V. Kühn, R. Harradi. and W. Hardt, "Expert System for Adaptive Flight Missions," Chemnitzer Informatik-Berichte, Chemnitz, Germany, June 2019

[4] M. Stephan, B. Battseren and U. Tudevdagva, "Autonomous Unmanned Aerial Vehicle Development: MAVLink Abstraction Layer," International Symposium on Computer Science, Computer Engineering and Educational Technology (ISCSET-2020), Lauta Germany, pp. 45-49, Oct 2020

[5] B. Battseren, U. Tudevdagva, W. Hardt and A. Banerjee, "Image Processing Based High Voltage Transmission Line Insulator Fault Detection," Proceedings of 13th International Forum on Strategic Technology (IFOST 2018), Harbin, China, pp. 955-960, June 2018

[6] M. S. Harras, "Detection of Physical Damages of High Voltage Power Line Insulator by Image Processing," M.S Thesis, Faculty of Computer Science, TU Chemnitz, 2020

[7] U. Tudevdagva, B. Battseren, W. Hardt and G. V. Troshina, "Image Processing Based Insulator Fault Detection Method," 2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), Novosibirsk, Russia, pp. 579-583, Nov 2018

[8] T. Jabid and M. Z. Uddin, "Rotation invariant power line insulator detection using local directional pattern and support vector machine," 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET), 2016, pp. 1-4, doi: 10.1109/ICISET.2016.7856522.

[9] S. Saleh, C. Rellan, S. P. Surana, J. Nine and W. Hardt, "Collision Warning Based on Multi-Object Detection and Distance Estimation," International Symposium on Computer Science, Computer Engineering and Educational Technology (ISCSET-2020), Lauta Germany, pp. 68-72, 2020

[10] S. Saleh, H. Saleh, M. A. Nazari and W. Hardt, "Outdoor Navigation for Visually Impaired based on Deep Learning," Proceedings of the 6th International Conference Actual Problems of System and Software Engineering, pp. 397-406, Nov 2019

[11] Ch. Surenjav, D. Byambaa, and U. Tudevdagva, "Deep-Learning-Based Insulator Detection Algorithm Study," Proceedings of MMT 2020 conference, Ulaanbaatar, Mongolia, pp. 28-32, 2020

[12] A. J. Chaudhry, "Burn-Mark Detection Based on Active Deep Learning" M.S Thesis, Faculty of Computer Science, TU Chemnitz, 2021

[13] A. Naeem and S. Peter, "Real-Time On-Board Deep Learning Fault Detection for Autonomous UAV Inspections" Electronics, vol. 10, pp. 1091, 2021

[14] A. Bochkovskiy, C.Y. Wang and H.Y.M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," ArXiV, arXiv:2004.10934, 2020

[15] M. W. Adou, H. Xu and G. Chen, "Insulator Faults Detection Based on Deep Learning," 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), 2019, pp. 173-177, doi: 10.1109/ICASID.2019.8925094.

[16] Z. Wang, X. Liu, H. Peng, L. Zheng, J. Gao and Y. Bao, "Railway Insulator Detection Based on Adaptive Cascaded Convolutional Neural Network," in IEEE Access, vol. 9, pp. 115676-115686, 2021, doi: 10.1109/ACCESS.2021.3105419.

[17] D. Sadykova, D. Pernebayeva, M. Bagheri and A. James, "IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging," in IEEE Transactions on Power Delivery, vol. 35, no. 3, pp. 1599-1601, June 2020, doi: 10.1109/TPWRD.2019.2944741.

[18] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang and D. Xu, "Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 4, pp. 1486-1498, April 2020

[19] Y. Wang, J. Wang, F. Gao, P. Hu, L. Xi, J. Zhang, Y. Yu, J. Xue and J. Li, "Detection and Recognition for Fault Insulator Based on Deep Learning," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1-6, 2018

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," ArXiV, arXiv:1506.01497v3, Jan 2016

[21] "Jetson Benchmarks," Accessed on: Aug. 10, 2021. [Online]. Available: https://developer.nvidia.com/embedded/jetson-benchmarks

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, and A.C. Berg, "SSD: Single Shot MultiBox Detector," European Conference on Computer Vision (ECCV 2016), Springer International Publishing, pp. 21-37, 2016

[23] P. Warden, How many images do you need to train a neural network?, Dec 14, 2017, Accessed on: June. 2, 2021. [Online]. Available: https://petewarden.com/2017/12/14/how-many-images-do-you-need-to-train-a-neural-network/

[24] "Jetson Nano: Deep Learning Inference Benchmarks" Accessed on: June. 2, 2021. [Online]. Available: https://developer.nvidia.com/embedded/jetson-nano-dl-inference- benchmarks

[25] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," ArXiv, arXiv:1804.02767, Apr. 2018, Accessed: Aug. 26, 2021. [Online]. Available: http://arxiv.org/abs/1804.02767

[26] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved YOLOv4-tiny," ArXiv, arXiv:2011.04244, Dec. 2020, Accessed: Sep. 14, 2021. [Online]. Available: http://arxiv.org/abs/2011.04244

[27] B. Hayes-Roth, "A Blackboard Architecture for Control," Artificial intelligence, vol.26, pp. 251-321, 1985

Technology and Access to Education During COVID-19

Bayartsetseg Batjav
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
bayartsetseg@mier.mn

Myagmarjav Usny-Ekh
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
myagmarjav@mier.mn

Sanjaabadam Sed
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
Sanjaabadam@mier.mn

Amarzaya Gulguu
Research and cooperation office
Khovd state university
Khovd, Mongolia
gulgooamarzaya@gmail.com

***Abstract*— Technological advances have had a significant impact on solving the problems of access to education. In this age, when information technology has been developing very rapidly, teachers and students of all levels are required to learn to work more effectively in the internet environment to process information, to achieve results and to apply their knowledge broadly in their work and life.**

Due to the outbreak of COVID-19 around the world, since the second half of the academic year of 2019-2020, all levels of educational institutions in Mongolia have been organizing their training activities online or in electronic form, which forces them to work in a new school environment in terms of both school management-organization and the activities by teachers and students. Mongolian government, the Ministry of Education and schools of all levels have taken continuous steps and issued orders and decisions which are directed at coordinating the activities of educational institutions and ensuring the quality of education during a high level of disaster preparedness in order to solve the challenges facing the higher education sector. In addition, it implements the policy recommendations and impact analysis of the COVID-19 pandemic made by UNESCO. Therefore, in this paper, we aim to provide the public with an overview of the impact of access to information technology on Mongolian education system during the pandemic, the required learning environment, the e-learning process, and the challenges of e-learning.

Keywords— e-learning, teacher skills, accessibility, pandemic

I. Introduction

In order to protect the interests of thousands of students during the pandemic (COVID-19) and ensure proper education, the order of the Minister of Education, Culture and Science No. A/43 of 2020 "On taking urgent organizational measures " was issued. And in connection with the decision by the Government of Mongolia and State Emergency Commission to suspend the activities of kindergartens, schools and public activities from January 27, 2020 to March 30, 2020 (again postponed until April 30, 2020), all levels of education were decided to be organized electronically. However, due to the lack of tools required for e-learning, as well as the lack of the availability and ability to use electronic devices, students of all levels are lagging behind.

According to Mongolia's 2020-2021 academic year statistics, there are 263,333 children in pre-school education institutions, 640,449 students in general education schools, 148,446 students in universities and colleges, and 37,806 students in technical and vocational education institutions.

Based on the results of the 2020 population and housing census, the living standards survey, and the UNESCO countries' distance learning readiness survey, it is estimated that 52.8–54.5 percent of secondary school students do not have access to e-learning due to lack of fixed internet access, 7.5–8.3 percent can't attend TV lessons because they don't have a TV, and 3.2 percent do not have access to any of remote teaching because they do not have access to electricity [1].

In addition, 75.0 percent of students who cannot attend distance learning due to lack of electricity, 68.7-69.8 percent due to lack of TV, and 40.0-42.0 percent of students due to lack of internet are herders' children.

According to a joint survey on distance learning by education institutions conducted in the academic year of 2019-2020, it is estimated that 46.0 percent of students regularly watched TV lessons, and 54.0 percent of students did not watch TV lessons regularly or did not watch them well.

II. Research Methodology

This research used both qualitative and quantitative methods and secondary sources. The sources include the results of online surveys taken from students and parents from both secondary school and universities, and other relevant reports that were conducted during the cOVID-19. The qualitative data supports the quantitative data analysis and results.

III. Results and Discussion

A. In terms of information technology access

Globally, advances and changes in information technology are leading to reforms in the education sector and the learning process, which necessitates the constant updating of the education system.

The percentage of cable TV users in Mongolia has been relatively stable over the past three years, due to the pandemic, with the number of Internet users at home increasing 18.4 percent in 2020 compared to 2019. This increase is due to the fact that educational institutions have switched to e-learning and are now offering tele-lessons for students because of the pandemic. As of 2020, the number of mobile phones per capita is 1.09, which indicates a high level of mobile phone use in our country (Fig.1.)

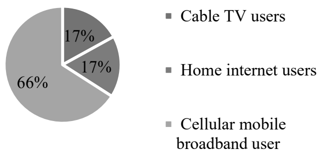


Fig. 1. Percentage of information technology users in the population (by year)

The percentage of cable TV users in Ulaanbaatar decreased by 3.8 percent in 2020 compared to the previous year while it increased in other regions and the percentage of home internet users is relatively higher in Ulaanbaatar than in other regions while being the lowest in the western region. As for the mobile communication users, it is the highest in Ulaanbaatar, with over 80.0 percent in the western and Khangai regions.

In order to increase access to information technology at the secondary school level, 25,000 laptops were given to teachers within the framework of the "Right Mongolian Child" national program, and 11,240 laptops were given to the computer training cabinets of secondary schools. However, in the academic year of 2018-2019, only 40.5 percent of all schools have adequate information technology cabinets. Only 48.2 percent of public schools and 6.1 percent of private schools have IT classrooms. 68.8 percent of soum schools do not have the cabinets dedicated to information technology. As of 2017, there were 572,752 students in secondary schools and the number of computers used for educational activities was 458.36. As of 2017, this figure was 12.5. In the academic year of 2018-2019, there were 30,411 full-time teachers in secondary schools, and the number of computers used for teacher development and teacher needs is 20,346. The national average number of teachers per computer is 1.5.

In the academic year of 2018-2019, there were 803 schools nationwide, of which 683 or 85.0 percent were connected to the Internet on the national average. On average, 94.3 percent of Ulaanbaatar schools, 88.8 percent of aimag center schools, and 77.5 percent of soum schools were connected to the Internet. This is considered to be the basic condition for conducting the training activities at home based on e-learning in the academic year of 2020-2021.

This suggests that the availability and use of information technology for e-learning is insufficient.

B. In terms of information technology use

At the general education level: Within the framework of a comprehensive plan to compensate for the learning delays of Mongolian secondary school students, 5284 TV lessons have been prepared and delivered to students by TV with sign language translation and 104 interactive lessons, 2775 e-lessons, 50 e-content, 30 radio and audio lessons have been prepared to be delivered through e-content and other websites.

Although tele- and e-learning lessons are provided to students, 172,300 students in 330 soums and 6,277 in 46 baghs, for a total of 178,577 students, are unable to attend tele- and e-learning due to a lack of infrastructure, internet, electricity, and family facilities [2].

It is reported that "There are a lot of students who don't have a TV at home, who live far in the countryside where there is no network, who are unable to watch TV at the same time because there are many students in one home, who don't have access to the internet at home, who can't afford to install or pay for data, or who don't have a smartphone. Even if it is possible to watch TV lessons, there are many families who do not have internet when they want to have homework to be checked by their teachers and to send their assignments. In addition, there are many students who are not able to keep up with the pace of the tele-lessons, who are unable to reverse-stop-catch-up, and in general, the content of the tele-lessons was at a pace for good students" [3].

When the main reasons for students not being able to attend TV lessons are classified, it was surveyed that 13.1 percent of students do not have a TV set, 33.25 percent do have a TV set but cannot watch e-lessons, 25.5 percent live in remote areas, and 11.7 percent do not have electricity.

According to the results of the "Teachers' Information and Communication Technology Readiness Survey", the self-assessment of primary school teachers' information and communication technology skills on a scale of 1–7 was rated at an average of 3.64-5.66. Teachers rated their ability to search and use materials and tools for education and training on the Internet as the best, and their ability to instruct and assist students in conducting research online as the worst. 41.0 percent of primary school teachers use information and communication technology from 10.0 to 30.0 percent in their regular classes.

At the tertiary level: In order to create an e-learning process and environs in universities and colleges, first of all, it is necessary to choose the right software management system and the methods of teaching. Nowadays, e-learning software, which is a key element of e-learning environ, open sources such as Moodle, C anvas, koha, Marc21, Model View Coltroller, Ruby, Rails, Lightweight Directory Access Protocol are being used globally. In addition, some universities with "Distance Learning Centers" (for example, MSUE) have faced difficulty in transitioning to full e-learning system. In order to change the teaching methodology, "Open Education Center" is gradually organizing the transfer of light bulbs, AR, VR, and MR technology to electronic form in the MOOC studio within the framework of the "Digital Professor" program [4].

According to the survey of e-learning activities of 79 universities, the student enrollment in e-learning is 90-100 percent in 24 universities, 80-89 percent in 29 universities, 70-79 percent in 14 universities, 60-69 percent in 13 universities and 0-59 percent in 6 universities. This is shown in the form of the ownership or property type (state or private) of the universities.

C. Teacher readiness for e-learning

We collected and compiled the data by conducting an open online one-month survey/ questionnaires on the supply of equipment to be used by the teachers of all levels of educational institutions for organizong e-learning, and on the forms and tools of training. 82.0 percent of the surveyed teachers are women, 18.0 percent are men and 31.2 percent work in pre-school education, 45.0 percent in general education and 23.8 percent in higher education.

TABLE I. CONDITIONS FOR TEACHERS TO WORK ONLINE

Working environs	Supply of training equipment	Ability to work in an Internet environment
Very good	103	112
	12%	13.1%
Good	416	403
	48.4%	47%
Bad	296	273
	34.5%	31.8%
Very bad	34	54
	3.9%	6.3%
Don't know	9	16
	1.05%	1.9%

60.4 percent of the surveyed teachers rated the supply of teaching equipment as very good and good, and 37.9 percent rated it as bad and very bad. 60.1 percent said they could work on the Internet and 38.09 percent said they could not. About 40 percent of the teachers faced these two problems when the organized e-learning.

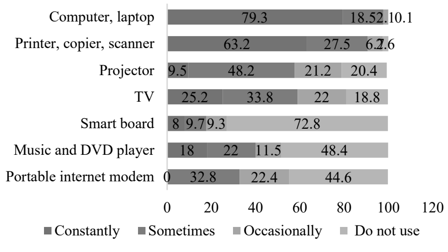


Fig. 2. How teacher use equipment (in percent)

The survey shows that the teachers use computers, printers and copiers more than other teaching techniques and equipment, while there are fewer teachers who use smart boards, televisions, music, DVD players, portable internet and modems. E-learning is conducted by using computers and laptops. Teachers who do not use computers or laptops use mobile phones and tablets.

83.8 percent of teachers surveyed said they use software such as Moodle, Zoom, MS Teams, and Google classrooms to make e-learning more effective, while radio training is less effective. As for the printed materials, 55.9 percent of the teachers surveyed said that they were effective, which is related to pre-school, primary and secondary education. When organizing e-learning at home, it is clear that both a teacher and a student are learning the knowledge and skills to work on the program.

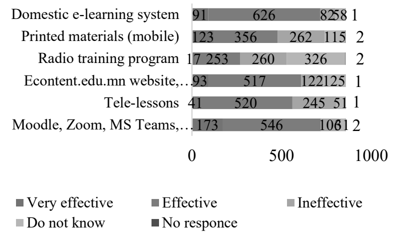


Fig. 3. Teachers' responses to distance and e-learning forms and tools

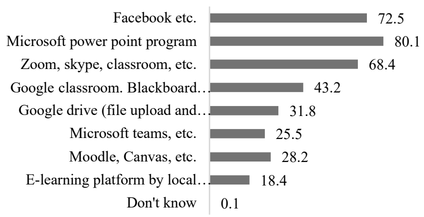


Fig. 4. Platforms and software used more in distance and e-learning

In the academic year of 2020-2021, more than half of the teachers surveyed said they use social networking sites such as Facebook, Zoom and Skype, and 68.4 percent use free e-learning platforms such as Google classroom.

The surveyed teachers use Microsoft Power Point to prepare their lessons and teaching materials, and 31.8 percent of the teachers use Google drive to upload and distribute the teaching materials.

42.7 percent of the surveyed teachers believe that the knowledge, skills and attitudes of teachers are suitable for organizing e-learning, 48.3 percent think that they are trying to keep up with technical and technological progress, and 38.4 percent think that they are lagging behind if they do not constantly develop themselves. This reflects the need for continuous development of teachers' knowledge and skills in conducting e-learning. In addition, about 88.6 percent of the teachers surveyed said that the support of the parents and guardians of the students is important in organizing e-learning.

In order to organize e-learning, the supports such as providing teachers with technical equipment for e-learning (65.3), improving computer and information technology knowledge and skills (58.2), cooperating with and learning from information technology teachers (78.5), improving e-learning environment (51.3) and strengthening school management capacity building for continuous professional development of school-based teachers (40.9) need to be provided. It is also important to provide teachers with the necessary training equipment for e-learning.

E-learning materials: 64.2 percent of the surveyed students said that they use mobile phones in e-learning. However, 53.0 percent of the parents surveyed said that the availability of technical equipment was insufficient for e-learning, and 60.0 percent of them said that their children use mobile phones for e-learning. This also shows that mobile phone is the most widely used information technology tool in our country. Therefore, it is necessary to develop and implement a unified policy on the introduction of other information technology tools.

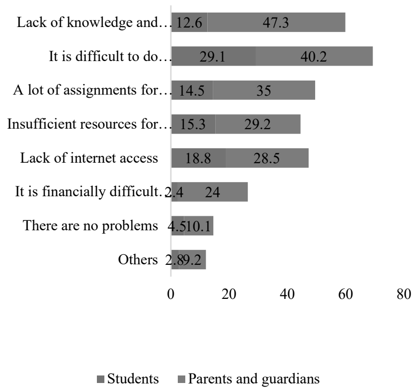


Fig. 5. Challenges of e-learning (according to students and parents)

Regarding the difficulties faced by university students in accessing e-learning, 29.1 percent of students surveyed said that it is difficult to complete homework and research work without a teacher's explanation and advice, and 18.8 percent said that they are lack of internet. The highest percentage of parents and guardians said that they do not have enough knowledge of e-learning technology and that it is difficult to do homework and research without a teacher's explanation and advice.

IV. Conclusion

The following conclusions are drawn from the results of questionnaires and surveys of the documents on the impact of e-learning environment, technological differences, and teachers' skills on e-learning access.

1. Lack of access to technology deprives people of equal access to education and limits access.
2. The lack of e-learning for children from low-income and remote areas indicates the weakness of e-learning policies.
3. More than 60.0 percent of students use only mobile phones in e-learning, which negatively affects the quality of education. Therefore, there is a need to introduce the use of other information technology tools in the state policy.
4. There is a need to systematically develop teachers' knowledge and skills in organizing e-learning and using platforms and software. In order to do so, there is a need to strengthen management capacity, such as continuous teacher development in the workplace, learning from other teachers, formalization of opportunities to share knowledge and skills with other teachers, and assessment of teachers' work.
5. Lack of responsibilities and support by parents and guardians in e-learning during the pandemic also has a negative impact on the availability of e-learning. Therefore, there is a need to develop and implement a step-by-step action plan that will have a positive impact on social psychology.
6. E-learning programs have only one option, which limits the learner's ability to master the content. Therefore, the issue of developing and delivering the programs with several versions is raised.

Mongolia's "Vision 2050", long-term development policy, includes the objectives for establishment of an open education system, the development of an integrated e-learning and distance learning platform, the full transfer of all levels of education to e-learning, and the development and dissemination of e-learning programs and content for students and citizens of all ages, and the improvement of teachers' skills for using information and communication technologies in training, for conducting e-learning and distance learning, and for bringing the content and standards of information and communication technology training to the international level to provide e-literacy education to students.

References

[1] Erdenechimeg E, "Education and Covid-19" Org. Res. 2020 [Online]. Available: https://www.kas.de

[2] "E-learning process for secondary school students," Mongolian Inst. Educ., Ulaanbaatar, Mongolia, Rep. 2021

[3] "Tele-learning during emergency, its results and suitability study", Civil Soc. Org. for Public Educ., Ulaanbaatar, Mongolia, Rep. 2021

[4] "Situation in the higher education sector of Mongolia during COVID-19 pandemic and opportunities to solve them", Ministry of Educ., Culture and Sci., Ulaanbaatar, Mongolia, Rep. 2021

[5] "Impact Analysis and Policy Recommendations on Higher Education during the COVID-19 pandemic," iesalc.unesco.org https://www.iesalc.unesco.org/en/2020/03/09/coronavirus-covid-19-and-higher-education-impact-and-recommendations/ (accessed 2020)

[6] "Primary Education sector analysis in Mongolia", Educ. and Soc. Develop. Cent. Mongolia, Ulaanbaatar, Mongolia, *Rep.* 2019

[7] "Recommendations for Higher Education during COVID-19 pandemic", WB, 2020

[8] "Research and recommendations on teacher information and communication technology readiness", Ministry of Educ., Culture and Sci., Ulaanbaatar, Mongolia, Rep. 2020

[9] "Results of the evaluation of the spring semester of the 2020-2021 academic year", National Univ. Mongolia, Ulaanbaatar, Mongolia, Rep. 2021.

[10] "Status on the Implementation of e-learning in universities" , Ministry of Educ., Culture and Sci., Ulaanbaatar, Mongolia, Rep. 2021

[11] "Tertiary Education and COVID-19: Impact and Mitigation Strategies in Europe and Central Asia," Tech. Rep. 2020. [Online]. Available: https://www.researchgate.net/publication/341781394_Tertiary_Education_and_COVID-19_Impact_and_Mitigation_Strategies_in_Europe_and_Central_Asia

[12] "The COVID-19 Crisis Response: Supporting tertiary education for continuity, adaptation, and innovation," openknowledge.worldbank.org https://openknowledge.worldbank.org/handle/10986/34571?show=full (accessed 2020)

Implementation of Distance Education Initiative Within Mongolian Higher Education Institutions During COVID-19

Buyandelger Banzragch
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
buyandelger@mier.mn

Lkhamtseren Bold
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
lkhamtseren@mier.mn

Gerelmaa Baljinnyam
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
gerelmaa@mier.mn

Abstract: The transition of the Mongolian Higher Education system to an online format since the 2019 – 2020 Spring semester as a result of the COVID-19 pandemic have placed a significant burden on instructors, students, staff and administrators, forcing them to adapt to the new realities. At the same time, there is significant pressure on Higher Education Institutions to continue to operate educational activities without interruption.

The purpose of this research is to analyze the implementation of distance education initiatives in Mongolian Higher Education Institutions during Covid-19.

The implementation of distance education initiatives were assessed in accordance with the following indicators:

- ***Distance education legal environment***
- ***Distance education format and technology***
- ***Student conditions during COVID-19***
- ***Instructor conditions during COVID-19***
- ***Assessment and diagnosis to overcome challenges presented by COVID-19 implemented and carried out at the institutional level***

Keywords: Higher Education Institution , distance education, policy, Information Technology, legal frameworks, platform, online content, internet.

I. Introduction

As a result of the closure of Higher Education Institutions in 175 countries, over 220 million students (13% of all students) experienced interruptions in education with some even abandoning their educational pursuits and Mongolian students are no exception.

There are three ways of looking at the current situation. Firstly, COVID-19 expedited the transition to distance education initiatives, bringing about a golden age for technology users and proponents All educational activities such as teaching, graduation and entrance exams have transitioned to an electronic format.

Mongolian Higher Education Institutions have undertaken activities to meet the infrastructure, technical, and equipment requirements of students and instructors in order to ensure high-quality, equitable distance education initiatives.

Secondly, the pandemic can be seen as an extreme but rare occurrence that presents an opportunity to assess the preparedness of Higher Education Institutions for emergencies. Higher Education Institution began developing and implementing Emergency Management plans and strategic plans.

Thirdly, there is an opportunity to undertake an accurate assessment of Higher Education under the current circumstances.

II. Distance Education Legal Environment

Based on the Mongolian Government legislation clause 24.2 of article 24, clause 10.4.4 of article 10 of the Emergency Management legislation, National Emergency committee directive 02 of February 26, 2020, Directive on the “Extension of the state of advanced preparedness” of February 19^{th}, and Directive A/43 of the Minister of Education and Science, 92 Mongolian Higher education institutions transitioned to distance learning initiatives in order to ensure the continuity of education.

The Higher Education legislation, Mongolian Government directive 139 on “Preventive actions to be taken against COVID-19 within the education sector”, Minister of Education and Science directive A/195 on “Measures to be taken within Higher Education Institution and college teaching activities” serve as the main regulations governing Higher Education during COVID-19. Higher Education Institutions are developing their own policies and regulations based on these decisions and directives.

Instructors, students and staff have also been also provided with advice and recommendations on their conduct during distance education initiatives and activities to develop distance education standards.

The National University of Mongolia, Mongolian State University of Medical Science, Mongolian State University of Education, Mongolian University of Science and Technology, University of Finance and Economics, Institute of Engineering and Technology, New Mongolian Institute of Technology and the Mongolian-German Institute of Technology have developed in-depth requirements on distance education classes in addition to general distance education policies that have proven to be helpful in the class development process.

Distance education software manuals, troubleshooting recommendations for staff and students were placed on school websites in video and text formats.

In addition to bachelor's, master's and doctorate education activities, student services and instructor support activities are also being carried out in an online format in accordance with internal regulations.

For example:

- Within the framework of the "Incorporation of innovative teaching methodologies based on advanced technologies, increasing the effectiveness of online and open education technologies and its use" objective of its Strategic Plan, the National University of Mongolia adopted their distance education class procedures on the basis of President's Directive A/89 of April 20 in order to coordinate the preparation and implementation of online classes. There procedures set clear requirements on the preparation of lecture and seminar materials as well as policies on the delivery of e-class materials to students and programs to be utilized. In addition to the policy, suggestions, manuals, explainer videos and program manuals were developed and distributed through social network and school websites.
- The Mongolian National University of Medical Science incorporated 31 clauses and 59 points on distance education into Chapter 4 of their Educational Technology and Quality Assurance policy in order to ensure the coordination of distance education classes, establish the requirements on programs, online learning resources as well as audio and videos used in distance education classes, coordination of distance education evaluation and students assessment as well as instructor rights and responsibilities and to ensure proper attention is paid to privacy and copyright concerns. The Mongolian University of Science and Technology and the Mongolian State University of Education had developed requirements on distance education classes in 2011 and 2017 respectively, leading to the majority of instructors being prepared for distance education classes and the conditions for a seamless transition to distance education were present.
- The Mongolian State University of Education approved the "Academic operation plan during lockdown" that encompasses a redistribution of department workloads in order to ensure equal distribution of responsibilities during lockdowns.

Additionally, Higher Education Institutions have adjusted policies and enacted temporary procedures in order to coordinate formative and summative assessments.

III. Distance Education Format, Software

As a result of COVID-19, Higher Education Institutions have transitioned their teaching, research, and scientific activities to an online format. Distance education activities are implemented in the following manners:

- Asynchronous (students undertake class activities in their own time). In this instance, e-classes, online resources, suggestions, and additional resources are placed on the school learning management system.
- Synchronous (students participate in classes at the same time). Programs like Zoom, Google Meet, Microsoft Teams, DingTalk are used.
- Hybrid (combination of asynchronous and synchronous formats).

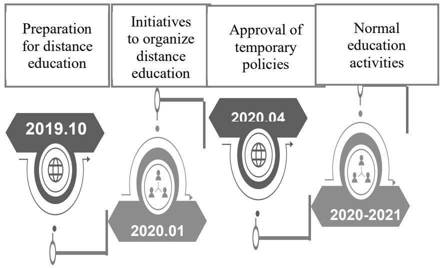


Fig. 1. Distance education timeline

A. Learning Management Systems

Public universities like the Mongolian University of Science and Technology (UNILMS), National University of Mongolia (SiSi) and the Mongolian State University of Education (Internal Learning Management System) are using a number of different Learning Management Systems while institutions such as Ulaanbaatar University, Global Leadership University, Institute of Engineering and Technology, and the New Mongolian Institute of Engineering and Technology are using a domestically developed x-Cloud system called Unicorn. However, the majority of Higher Educational Institutions are using open platforms such as Google, Moodle, Neolms as Learning Management Systems during their distance education operations. Most institutions are using programs such as in delivering their classes to domestic students and VOOV to students abroad. Some institutions are even widely using social networking platforms such as YouTube and Facebook.

Learning Management Systems encompass an instructor and student dashboard, administration dashboard, e-class lecture, seminar and laboratory dashboard, additional resources section, student discussion and assessment dashboard and a student attendance statistics dashboard.

B. Distance Education Software

In addition to Learning Management Systems it is estimated that Higher Educational Institutions have procured licenses for and utilize an average of 10-18 programs and apps to develop classes, record and edit classes and record screens, audio.

For example, programs such as Camtasia, AutoEffect, Illustrator, Adobe, Audacity, and Asciicode are utilized for distance education initiatives. There are even some instructors referring to platforms such as www.edx.org, www.coursera.com, www.udacity.com, and www.mooc.org in preparing and delivering their classes. Following the transition to distance education classes, school websites, branch school, department, and unit Facebook pages have received significantly more traffic.

Most Higher Education Institutions have developed distance education methodology workshops (organizing distance education lectures and seminars, preparing online resources and manuals) with explainer videos, manuals and suggestions also being delivered to instructors through e-mail.

C. Distance education departments and units

Larger public universities have formed distance education departments and units, making relevant organizational and structural adjustments. For example, the National University of Mongolia and the Mongolian State University of Medical Sciences have expanded their Instructor Development Centers into Instructor Development and Distance Education Departments. The National University of Mongolia Instructor Development and Distance Education Department established a “Digital Professor” studio with lightboards for instructors to prepare distance education content and implements comprehensive New Faculty Orientation programs.

The Mongolian State University of Medical Science Instructor Development and Distance Education Department organizes the “Best Distance Education Lecture” competition and “Digital Instructor” workshop in order to enhance distance education classes and content quality and provide opportunities for instructors to develop video class content in a multimedia studio.

Established in 2017, the Mongolian State University of Education Distance Education Center not only plays a significant role in the implementation of distance education institutions within the university but also organizes national and international scientific conferences.

The Mongolian University of Science and Technology Open Education Center has prepared 363 classes in disciplines such as Mathematics, Physics, Chemistry, Material Resistance, Theoretical Mechanics, and Vehicle Components in a lightboard studio using lightboard, AR, VR, and MR technologies in order to reform teaching methodology. Moreover, instructors are also provided with the opportunity to be encompassed in “Distance education challenges, experiences and opportunities during COVID-19” as well as certificate courses on Coursera.

IV. Student Conditions During Covid-19

During the 2020-2021 academic year, 147,300 students were encompassed in 88 Higher Education Institutions, 24,900 of whom were new students.

As comprehensive government policies and initiatives have been directed towards high-quality distance education classes, new students faced little challenge in participating in distance education initiatives.

TABLE 1. HIGHER EDUCATION STUDENTS

Indicator	2019-2020	2020-2021
1. Number of Higher Education Institutions	92	88
Public	21	20
Private	71	65
2. Number of students	148446	147293
Female	90573	89463
3. Students attending public institutions	82901	76844
Female	49238	44738
4. Students attending private institutions	65355	63545
Female	41212	40430

Source: www.1212.mn National Statistics Database

A Mongolian Institute for Education Research and Mongolian State University of Education joint study found that of 21,417 respondents, 64.2% used mobile phones to participate in classes while 29.6% used laptops to participate in classes. If we delve into some of the challenges in distance education mentioned in the study, 48.3% of respondents stated that their workload increased, 20.1% stated pointed at internet and connectivity issues, 18.9% stated that there was a lack of open online resources, 23.2% mentioned a lack access to the instructor, 13.8% pointed to difficulties in carrying out independent work without support from the instructor and 29.1% found it challenging to navigate the platforms used in classes.

TABLE 2. ADVANTAGES OF DISTANCE LEARNING

№	Answer	Number of frequencies	Percentage
1	Reduction of the potential for infection	16741	27.8
2	Independent learning	6877	11.4
3	Studying in accordance with their own learning style	3685	6.1
4	Using advanced educational technologies	2037	3.4
5	Enhanced quality of learning	516	0.9
6	Ability to rewatch lectures, videos and other materials	8762	14.6
7	Ability to download lectures, videos and other materials	4756	7.9
8	Ability to challenge oneself	1947	3.2
9	Economic savings	4476	7.4
10	Flexible schedule	5437	9.0
11	N/A	4887	8.1
12	Others	34	0.1
	Total	60155	100.0

As mentioned previously, 48.3% of students believe that their workload increased as a result of the implementation of distance education initiatives (Fig.1).

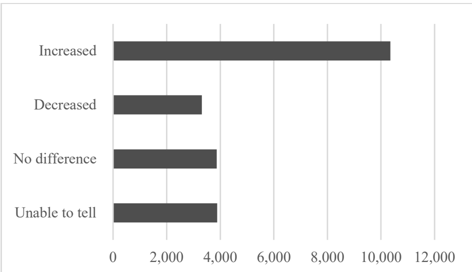


Fig. 1. Student Distance Education Workload

Higher Education Institutions are regurlary conduct assessments into learning gaps resulting from the shift to distance education initiatives throughout the course of the Spring semester of the 2019/20 academic year and the 2020/21 academic year and making relevant adjustments where necessary. For example, the Mongolian University of Science and Technology monitors and generates feedback on distance education activities throughout the course of the semester and at its end. When we compare student GPA's to that prior to the pandemic, there is no decrease and there are even instances of student GPA's increasing during distance education initiatives.

Clause 2.3 of the National University of Mongolia "Coordination of education activities" directive of January 25, 2021 holds for the assessment of distance education classes in the 3-4th week of classes and the suspension of classes deemed to not be meeting requirements. Within the framework of this clause, the Curriculum Quality Assurance Department carried out assessments of distance education classes between February 23rd and March 1st, concluding that the conditions for the effective implementation of distance education classes were present.

A Mongolian Institute for Education Research study found that over 20% of students did not participate in distance education classes. One of the primary reasons for this was connectivity challenges outside Ulaanbaatar and the high price of mobile data preventing students from downloading audio and video classes. It is also worth noting that distance education challenges also arose from the discrepancies in the technology available to students.

The following activities are implemented by Higher Education Institutions to support students during these times:

- Distance education class information, manuals, and suggestions were placed on Learning Management Systems in video and document formats;
- Distance education manuals were placed on school websites and distributed through other avenues;
- Information and suggestions on preparation for and participation in distance education classes were provided to students;
- Instructors, department heads and student service specialists undertook the responsibility to resolve and challenges arising from distance education initiatives in a timely and supportive manner;
- Higher Education Institutions are working with mobile carriers on initiatives such as obtaining free access to their Education Management System for students or offering students unlimited internet on affordable fixed data plans;
- Initiatives such as allowing students to pay in installments, increasing scholarships, and allowing students to postpone tuition payments until next year were undertaken.

In addition to conducting classes in a distance manner Higher Education Institutions, special attention was paid to initiatives to support the socio-economic condition of students. Some of those initiatives included:

- During the 2020-2021 academic year Higher Education institutions did not raise their tuition fee;
- Every Higher Education Institution is offering additional scholarships and financial support;
- Higher Education Institutions are conducting education activities despite delays in special funding and tuition payments;
- Dormitory fees were fully refunded to students with some instances of payments being credited to next year's fees at the request of students;
- Student that refused to participate in distance education classes were offered a full discount.

The National University of Mongolia Student Services Department developed a 4-part psychological recommendation and undertook activities to support student psyche during lockdowns as well as placing psychological recommendations on their site. (https://student.num.edu.mn/?page_id=6841).

V. Instructor Support Activities During COVID-19

During the 2020/21 academic year, there were 7,143 instructors and 11,970 staff were employed by the 88 instructors. This constituted a 2.5% reduction in instructors from the 2019/20 academic year despite a 6% increase in students (Fig. 2).

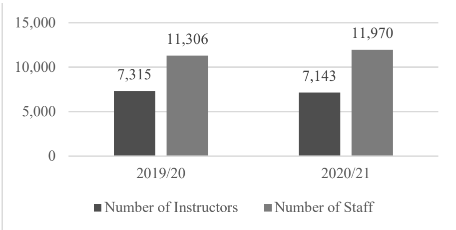


Fig. 2. Higher Education Sector Instructors and Staff

The following activities are examples of initiatives implemented by Higher Education Institutions in the 2019/20 academic year in order to support instructor development:

- Instructors were provided with the equipment required to conduct distance education initiatives such as computers, cameras, headphones, microphones, and televisions

- Distance education recording studios and the environment for effective distance education initiatives were provided to instructors

- Instructors were provided with 10Tb of storage space on the distance education on Learning Management Systems

- Instructor salaries were increased wherever possible

- Instructors and staff were regularly provided with health and sanitary materials

- National and international conferences were organized in a distance manner

- Instructors were provided with recommendations and suggestions on working with students and supporting them

- Distance education program workshops were offered to instructors and staff with relevant recommendations also being provided.

A number of initiatives were implemented to enhance instructor ICT and distance education methodology competency and improve teaching and learning activities. Some of those initiatives included:

- Software use workshops (Mongolian State University of Arts and Culture, Mongolian University of Science and Technology, National University of Mongolia and the Mongolian State University of Medical Sciences);

- Distance education platform use workshop (Mongolian University of Science and Technology);

- Principles of preparing digital content and its effective incorporation in teaching and learning workshop (Mongolian University of Science and Technology);

- What is distance education? Distance education content,

Open Edx, Microsoft Teams program use (National University of Mongolia);

- Preparing for distance education initiatives. Preparing distance education video content (National University of Mongolia);

- Downloading Teams Insights and other software, and its use workshop (National University of Mongolia);
- Teaching and Learning workshop (Higher Education Instructor Development Department of the Institute for Instructors' Professional Development);

- Distance education methodology workshop cluster (Mongolian State University of Education);

- Distance education theory and methodology (Mongolian State University of Arts and Culture).

VI. Initiatives to Overcome the Impact of Covid-19 Taken at the Institutional Level

Higher Educational Institutions are assessing and analyzing their distance education conditions. Some of these initiatives include assessing student achievement, conducting student and instructor satisfaction surveys, and monitoring distance education class participation as well as the implementation of classes.

Instructors and students encompassed in satisfaction surveys offered the following thoughts:

- Online system development and updates are required;
- Students are ranging difficulties ranging from a lack of proper learning environment and opportunities to collaborate with peers, inability to take part in seminars, technical issues as well as lack of access to additional resources leading to a reduction in;
- Learning management portals need to be improved with student and instructor collaboration dashboard and the conditions for effective individual and collaborative study need to be established;
- Instructors need to enhance the distance education content and update video classes with up-to-date examples, and ensure the conciseness of distance education classes;
- Instructors need to be provided with required technical equipment and financial support in order to enhance their working condition;
- There are difficulties originating from instructors using their own equipment for distance education classes;
- Distance education development needs to be based on the intricacies of each class;
- Additional funding for the distance education environment is required;
- There is a lack of resources and manuals to provide instructors support on distance education methodology.

The National University of Mongolia conducted two surveys within the 2020/21 academic year in order to assess the education conditions and quality. In the Fall semester, 92% of instructors assigned tasks to students in order to reinforce their knowledge and 85% of instructors offered feedback on these tasks. Student assignment completion grew from 58% in the Fall semester to 67.6% in the Spring semester. If we look at the student satisfaction surveys, 73% of instructors delivered more than 80% of their content while that number grew to 84.3% in the Spring semester. Moreover, 56 – 79.7% of class materials were considered to be above average.

Relevant adjustments and improvements are made based on the results of the student and instructor satisfaction surveys where appropriate.

VII. Conclusions

- Covid-19 impact mitigation strategies are implemented at the national and institutional levels and the legal environment for distance education initiatives is being established.

- Higher Education Institutions are implementing initiatives the socio-economic conditions of instructors, staff, and students in order to support the uninterrupted continuation of teaching and learning, research and science operations. For example, financial support initiatives included increased scholarship, grant, loans and flexible payment arrangements for students while methodological support included the organization of distance education methodology workshops as well as psychological support. Initiatives to support access to required technology and Learning Management System included enhancing internet speed and data space as well as the provision of required distance education software while health and sanitary initiatives included encompassing students and instructors in private medical insurance.

- Research into the Higher Education sector conducted during the pandemic indicated that there is a lack of information and transparency surrounding equality. There is an especially noticeable absence of policies to protect and support at-risk students

YIII. Suggestions and Reccomendations

National level

1. Attention needs to be paid to the development of a platform containing the best Higher Education sector content.

2. Development of open education principles into the Higher Education sector.
3. Announcing and organizing scientific and research competitions on distance education theory and methodology.
4. Development of quality assurance mechanisms for online content, open resources and other resources.
5. Attention needs to be paid to ensuring the provision of internet and connectivity, equipment, virtual laboratories and simulators.
6. Instructor ICT core competency standards need to be implemented and they need to be encouraged to continue developing high quality distance education materials and resources by the Ministry of Education and Science by organizing competitions or by tying the development of these resources to instructor evaluations, career advancement opportunities and financial incentives.
7. Provide support and loans to students in procuring adequate computers.
8. Mobile carriers need to research avenues to support students outside Ulaanbaatar in participating in distance education initiatives (free access to education management systems and other programs etc.).
9. The procurement of official licensing for online resources used by students is essential.
10. Initiatives to develop distance education programs need to be supported and incentivized.
11. Attention needs to be paid to copyright considerations when digitizing books and learning resources.

Higher Education Institution level

1. There is a need to educate instructors and students on distance education and redesign teaching and learning operations.
2. Student and staff distance education assessment needs to be reflected in Higher Education Institution policies. It is also essential for institutions to develop mechanisms to calculate distance education class preparation workload as well as to assess and award bonused for distance education classes.
3. Supporting and incentivizing research and projects surrounding hybrid teaching opportunities and innovative teaching methods is essential.
4. It is necessary to evaluate the possibilities for conducting research and experiments in a virtual manner.
5. Higher Education Institutions need to develop innovative methods for assessing student learning based on the intricacies of each program and constantly make relevant adjustments. Set up feedback mechanisms to collect and analyze data on course results and student responses, providing appropriate quality assurance of education technology and online delivery.
6. Higher Education Institutions need to revisit mechanisms to assess student activities, compile and analyze information on class statistics.
7. Programs for foreign students need to be developed. Activities and flexible programs to support the interests of international students including but not limited to tuition remission, living arrangements and improving living conditions is required.
8. Providing socio-economic and psychological support to at-risk students
9. Weaknesses in infrastructure need to be determined. Engage financial management and procurement teams as early as possible, to understand the opportunities and constraints related to purchasing of technology, licensing, hardware and software for students and academic staff, etc.
10. Student learning achievements need to be regularly analyzed and a constant feedback mechanism needs to be in place.
11. Within the framework of UNESCO's "No student left without attention" national principle, special attention needs to be paid to at-risk students and students with disabilities.

References

[1] Parliament, M. (2002, May 03). *Law of Education.* www.legalinfo.mn : https://legalinfo.mn/mn/detail/9020

[2] Parliament, M. (2002, May 03). *Law of Higher Education.* www.legalinfo.mn: https://legalinfo.mn/mn/detail/251

[3] Parliament, M. (2016, February 04). *Sustainable Development Principles of Mongolia-2030.* www.ilo.org: http://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---ilo-beijing/documents/publication/wcms_542890.pdf

[4] Parliament, M. (2010). *Roadmap for higher education reform 2010-2021.* www.num.edu.mn: https://www.num.edu.mn/files/ZHSHUA/DB%20shinechlelin%20zuraglal.pdf

[5] Mongolia, G. o. (2010, June 09). *Development of a Campus University.* www.num.edu.mn

[6] Science, M. o. (2019). *E-content database.* https://old.econtent.edu.mn

[7] Science, M. o. (2017). *Education Information System.* www.esis.edu.mn: https://www.esis.edu.mn/hr/index.html#

[8] Research, M. I. (2019, 12 31). *Country Backround Report.* Retrieved from www.meds.gov.mn: https://www.meds.gov.mn/post/67463

[9] UNESCO. (2019). *Mongolian Education Policy Analysis.* www.cdn.greensoft.mn: https://cdn.greensoft.mn/uploads/users/2649/files/PDF%20files/EPR%20final_web%20version1020.pdf

[10] Education, [. S. (2019). "Developing multimedia software and incorporating into education" 11th national competition . Ulaanbaatar: (unpublished)

[11] Munkhbaatar, M., Dunjinnamdag, O. (2017). *Incorporating ICT into education.* www.mmt.edu.mn/: http://mmt.edu.mn/static/upload/template/template29.pdf

[12] Altantuya, Y. (2018). *Feasibility study into the incorporation of an Information Management System into the education sector.* Ulaanbaatar: (unpublished)

[13] Mongolia, C. R. (2018). *ICT core indicators.* www.crc.gov.mn: https://crc.gov.mn/k/2ZV

[14] Mongolia, C. R. (2019). *ICT core indicators.* www.1212.mn: https://crc.gov.mn/k/32C

[15] IRIM, M. o. (2020). Tele-classes during emergency conditions, its results and conditions. Ulaanbaatar : (unpublished)

[16] Mongolia, N. U. (2020, November 18). *Comparative study of Covid-19 pandemia measures.* www.num.edu.mn: https://bs.num.edu.mn/wp-content/uploads/2020/12/Covid-study-by-NUM-202011.pdf

[17] Science, M. o. (2021). *Report of working group on the analysis of the Higher education Sector distance education learning environment,*

M&E, analysis of initiatives and development of future recommendations. Ulaanbaatar. (unpublished)

[18] MNUE, Mongolian Institute for Educational Research. (2021). *Instructor survey results.* Ulaanbaatar. (unpublished)

[19] Mongolia, N. U. (2021). 2020/21 academic year Fall semester education implementation evaluation result. Ulaanbaatar: (unpublished)

Design of Two-Stream Convolutional Neural Network for Vision-Based Obstacle Avoidance in Indoor Mobile Robots

Chaymae El Mechal
Electrical Engineering Department
Faculty of Science and Technology - University Sidi Mohamed Ben Abdellah
Fez, Morocco
chaymae.elmechal@usmba.ac.ma

Najiba El Amrani El Idrissi
Electrical Engineering Department
Faculty of Science and Technology - University Sidi Mohamed Ben Abdellah
Fez, Morocco
najiba.elamrani@usmba.ac.ma

Mostefa Mesbah
Department of Electrical and Computer Engineering
Sultan Qaboos University
Muscat Oman
m.mesbah@squ.edu.om

Chaimae Zai
Electrical Engineering Department
Faculty of Science and Technology - University Sidi Mohamed Ben Abdellah
Fez, Morocco
zaichaymae@gmail.com

***Abstract*—Obstacle avoidance in vision-based mobile robot experiences considerable challenges in its performance and reasoning. Avoiding collisions with obstacle are priorities in mobile robots. Its objective is to train the mobile robots how to navigate in an unknown indoor environment while avoiding colliding with objects. There are different new types of algorithms to train a mobile robot for this task. Recently, the Convolutional Neural Network (CNN) has been rapidly growing in robot navigation, image classification, object detection, and recognition. It provides an excellent performance that depends on the CNN architecture, hyper-parameters, and used data.**

In this work, the development of a CNN method is investigated and an RGBD architecture for obstacle avoidance in an indoor environment is proposed. Our architecture comprises two essential CNN processing streams - one designed for RGB images and the other for depth information to consecutively combine them with a late fusion network. The network takes RGB and Depth images as inputs to get one of seven (07) available control commands as outputs. A new dataset is collected using Kinect camera attached to a mobile robot (Turtlebot) which contains RGB and depth images while robot orientation data is obtained by an Inertial Measurement Unit (IMU) attached to the same robot. Several One-Stream and Two-Stream CNN architectures are trained using the collected data with different training options and hyper-parameters. The output accuracy is evaluated and recommendations are provided accordingly. The final results show that the accuracy of a Two-Stream CNN which RGB (93.3%) and depth (86.5%) inputs give better accuracy. Furthermore, the network performance can be improved by tuning the network's hyper-parameters and changing the number of layers. The CNN shows a great potential to achieve a high classification accuracy in avoiding obstacles for mobile robots.

Keywords*—*Robot vision; obstacle avoidance; image fusion; deep learning; convolutional neural network. Two-stream CNN.

I. INTRODUCTION

The contemporary technological advancements in the design and production of robotics have intensely focused on autonomy. Complete autonomous robots are designed with the ability to successfully navigate through the environment and complete the intended tasks [1]. One of the primary issues with navigating the environment is to avoid obstacles, using ranging sensors and algorithms, which collectively detect and analyze obstacles, then guide the robot safely in its movement. Robotics uses a wide range of sensors such as laser, sonar, depth images, and optical flow sensors. Most of these methods present various issues related to the resolution, complexity, cost, and data, inhibits proper obstacle avoidance in different environments [2]. To overcome these challenges, cameras are integrated with deep learning networks to capture the images in different environments and guide the robot safely [1], [3].

Convolution neural network is one type of deep machine learning that has been widely used for objects recognition and image classification. A variety of deep CNN algorithms have been reported by several groups who most of them took advantage of CNN's ability to extract features. However, most researchers and obstacle-avoidance algorithm developers for vision-based mobile robots encounter considerable challenges in computational efficiency as well as redundancy. The main question is how well the current techniques can work in real environments.

A CNN-based algorithm controls the robot through a vision system that captures RGB and depth images then assesses their location to direct the robot without collisions with obstacles [4]. CNN was used to enable mobile robots to walk around in unknown environments [5]. In particular, the study explored algorithms for mobile robots that used several CNN layers with the decision-making process. The entire system was trained to end by taking only RGBD information as the input and to produce sequence of main moving directions as the output so that the robot can autonomously explore the environment. The robot was trained and evaluated through real world environmental conditions. It has been reported that this technique brought new evolution into the field of image fusion for the robot control through a vision system that the obstacle avoidance successfully [5].

The techniques of fusion RGB and Depth have already been successfully used with CNNs to enhance the accuracy of the classification task [7]. The aim of this work is to use only the feature fusion, which means the convolutional layer takes the characteristics of RGB and Depth images before performing the fusion. For this purpose, we have chosen the Two-Stream CNN method. This technique utilizes two same parallel nets: one takes RGB images as input, while the other takes depth images. The features of both types of images will be extracted, then followed by fusion layer which uses both features (RGB and depth) before feeding them to the fully connected layer.

Most RGBD object detection methods use two parallel CNN streams for RGB modalities and depth information, which are then combined into early or late layers [8, 9, 10, 11, 12]. In this paper, we propose an algorithm which consists of two-stream to process depth and RGB images for achieving more information. By this way, one can have more features from two types of inputs. Each stream consists of a convolution layer, ReLU, and Max-pooling by investigating the need for a fully connected layer in each stream as well as the fully connected layer at the end. This method can successfully fuse RGB and depth information by developing the CNN algorithm. This method turned out more efficient than the one based on only RGB or depth images. The classifiers were optimized by training the network's final layers with data collected locally using image sensors attached to a TurtleBot mobile robot. The training of convolution neural networks requires a sizable aggregate (volume, mass) of training data. Thus, after testing the deep learning algorithm using the benchmark dataset, a new dataset will be built using the Turtlebot platform. An indoor depth data-sets from different scenarios will be collected. During the data collection process, an instructor will operate the mobile robot to explore an unknown indoor environment without obstacles smash. The synchronized RGBD images obtained by the Kinect camera and the angles acquired by the IMU will be recorded.

II. Convolutional Neural Networks

Keep A generic CNN architecture is a typically composed of diverse types of layers called building blocks such as convolution layers, pooling layers, and fully connected layers (Fig. 1). A brief representation was given for each layer separately below [13], [14].

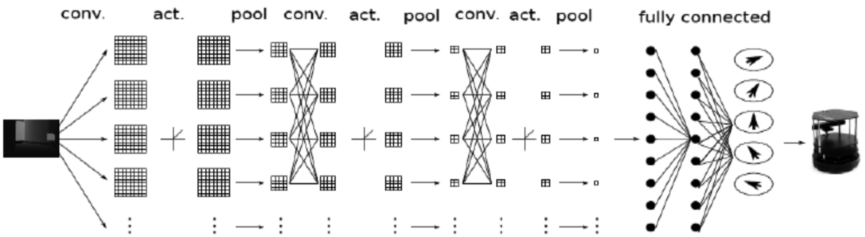


Fig. 1. Structure of the CNN [7]

A. Convolution layer

The Convolution layer (CL) is designed to take advantage of the 2D structure of learnable filters of input images (also called convolutional kernels) which are convolved with a given input to generate an output feature map. It executes most of the computational heavy load, this learning procedure involves a random initialization of the filter weights at the start of the training. The mathematical expression of the overall operation is given by:

where, H_{ijk} denotes the pixel value at coordinate (j, k) of the (i[th]) output feature map, and the convolutional layer convolves X with k filters $\{W_i\}$. b_i is the i[th] element of the bias vector associated with the i[th] convolution kernel (filter) and function f is a nonlinear activation function such as the ReLU by applying zero-padding through the frame axis before every convolution.

B. Activation Function

The activation functions convert the weighted sum of inputs into the artificial neurons, which is the abbreviation of Rectified Linear Units (ReLU). It also decreases the likelihood of a vanishing gradient problem. A ReLU layer applies a threshold operation to all of the values in the input volume using the function is defined as follows:

$$H_i^{\sim} = max(0, H_i)$$

in which H and H˜ are the input and output respectively. In particular, this layer just turns all the negative activations to zero.

C. Pooling operation

One of the essential tasks of pooling is to decrease the size of the output. This leads to a reduction of the number of the network parameters and, hence, an increase in the speed of the learning process. The most commonly used pooling techniques are the max-pooling and the average pooling.

D. Fully connected layer

Convolution and pooling layers act as an automatic feature extractor. Once the features are extracted, they are sent to one or more fully connected layers followed by a softmax function layer for classification.

A Pre-trained CNN is used as a feature extractor for training an image category classifier by using TS-CNN [15]. A higher performance is achieved by the Two Stream Convolutional Neural Network (TS-CNN) in comparison to other architectures. It is worthy to note that identity jump connections in residual blocks and improved detection of smaller objects in images are among the interesting properties of the TS-CNN architecture.

Convolution and pooling layers act as an automatic feature extractor. Once the features are extracted, they are sent to one or more fully connected layers followed by a softmax function layer for classification.

A Pre-trained CNN is used as a feature extractor for training an image category classifier by using TS-CNN [15]. A higher performance is achieved by the Two Stream Convolutional Neural Network (TS-CNN) in comparison to other architectures. It is worthy to note that identity jump connections in residual blocks and improved detection of smaller objects in images are among the interesting properties of the TS-CNN architecture.

III. PREPARATION OF DATASET AND PRE-PROCESSING

Data and corresponding RGB and depth images are used for classification. Microsoft Kinect , attached to the TurtleBot 2 Personal Robot [10] and running the Robot Operating System (ROS), is used to capture data from indoor scenes. The Turtlebot which is equipped with a mobile base, a laptop, and a Kinect sensor is controlled by a joystick for navigation in different paths. In this study, ROS is an operating system that runs on the controller unit and interacts with different components of the robot [16]. Our system generates RGBD images at a resolution of 640×480 pixels.

After processing and resizing the obtained images, they were classified into seven labels .The labels were chosen to improve the classes of work done in Ref [7]. In that work, there were 5 classes, so we increased them to 7, which is good enough for most robotic applications. Also, adding more classes improves the testing accuracy of the data. The choice of the range of angles was done by dividing the equal range of angles expect classes one, which is “go straight forward”. The maximum angle is 67.5° as in most of real applications.

The data was sampled and discretized into seven path labels which are “go-straightforward (1)”, “turning left with 22.5° (2)”, “turning- left with 45° (3)”, “turning- left with 67.5° (4)”, “turning -right with 22.5° (5)” turning -right with 45° (6)” and turning -right with 67.5° (7)”. This classification was done based on the reading of IMU sensor. The angles obtained from the IMU sensor were in a range of -30° to 30°. These angles are divided as follows:

TABLE I. CLASS DEFINITION BASED ON IMU READING

Angle (°)	**Class**
$-1 < \theta < 1$	(1) “go-straightforward”
$1° \leq \theta < 22.5°$	(2)” turning -right with 22.5 degree
$22.5° \leq \theta < 45°$	(3)” turning -right with 45 degree
$\theta° \geq 45$	(4)” turning -right with 67.5 degree
$-22.5° < \theta \leq -1°$	(5) “turning- left with 22.5 degree
$-45 < \theta \leq -22.5$	(6) “turning- left with 45 degree
$\theta \leq -45$	(7) “turning- left with 67.5 degree”

IV. STRUCTURE OF THE TWO-STREAM CNN MODEL

Few image fusion algorithms have been proposed within the framework of image classification based on two CNN streams to provide better performance. In CNN's two-stream structure, several important operations need to be conducted. Since there are two inputs to the CNN, the general idea is to train two independent CNNs where, one is handling the RGB images and the depth information as shown in Fig. 2. Then it combines the results of the two trainings and obtains a final recognition result.

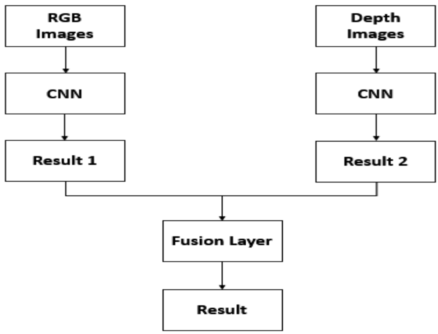


Fig. 2. The structure of the two-stream CNN model

V. EXPERIMENTS

The selection of a neural network structure is a challenging task as there are several criteria to be considered. There is no direct way to build the CNN model, so trials of many models and comparison of the network performance is the only way to get the best model for a particular data set. Therefore, we used the previous work as a starting step to build a new architecture that gives better accuracy [17]. The reported CNN model consists of three convolutional stages followed by one fully connected layer and one classification layer. Each convolutional stage consists of a convolution layer, ReLU layer, and Max-pooling layer.

The architecture implementation in Ref. [17] was studied thoroughly to determine the best possible setup for the learning task. However, the model was implemented for depth images only with 5 classes [17]. In this work, we increase our work on the categories of the images: RGB and depth, and we varied different parameters in order to reach a good learning. Thus, we have two CNN models types, namely (1) the one-stream model that takes one type of image as input, and (2) two-stream models that take both RGB and depth images as input. We made several modifications to the CNN architecture so that we could implement our CNN.

Finally, we used two streams that consist of four convolution layers and four fully connected layers. These two networks are then merged by a concatenation layer followed by two dense layers and an output layer. Figure 3 shows the adopted model. The network achieved in this work is a two-stream convolutional neural network for the detection of RGBD objects. The input to our network is a pair of RGB and depth images. In Fig. 3 each stream (blue, green) consists of five convolutional layers and two fully connected layers. The two streams converge in a fully connected layer and a softmax classifier (Grey).

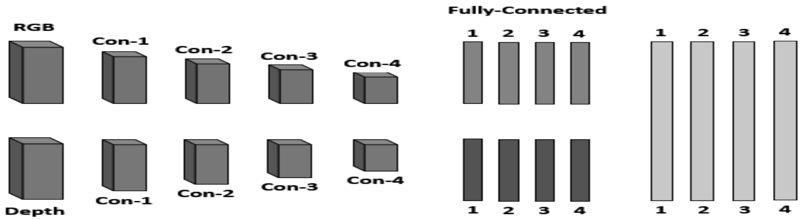


Fig. 3. Two stream CNN model adopted in our work

A. Experimental setup

Step 1: The images are scaled to get the proper image size, and all duplicated images for each scenario are removed. The size of the original images (640×480) are reduced to almost a quarter. Toward this objective, another pre-processing approach is designed: we reduced the longer side of the original image to 256 pixels, resulting in an image of 256 × N or N × 256 pixels. Next, we tile the edges of the longer side of the axis from the shorter side. We applied the same scaling operation for the other RGB and depth images and use them as inputs for the CNN network.

Step 2: In this step, we normalized all images between 0 and 255. Then we applied a jet colormap on the given image which transforms the input from a single channel image to a three-channel image. During the colorization procedure, the RGB and Depth image is used for each pixel of size W×H, because the colorization provides sufficient common structure between a depth image and an RGB image to learn the appropriate feature representations and improve the efficiency and effectiveness of data.

Step 3: At this stage, we have trained the combined RGB and depth images with different two-stream CNN architectures. This CNN contains two similar parallel networks: one takes RGB images as input while the second take depth images as input. The features of the two images are mixed by concatenating layer which are followed by another dense layer, the softmax layer. That means we have a dense layer of each network and another after mixing learning from both networks. Thus, we obtained a good result for the (RGBD) four-channel image.

The CNN is trained by changing the hyper-parameters for increasing the accuracy of the network. Setting the hyper-parameters needed experiences and a lot of trial for founding the suitable parameters for our dataset to give us better testing accuracy. After changing the number of parameters, we get different results for each model. For the proposed model, it was found that the best learning rate is 0.001 and the mini-batch is 32. The multi CPU core was used to reduce the training time by half, where the momentum was 0.9 and the weight decay was 0.0005.

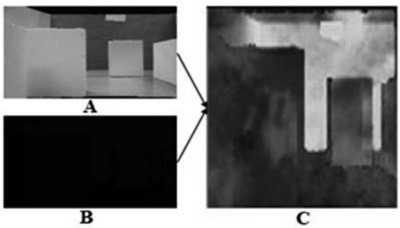


Fig. 4. A) RGB, B) Depth, C) RGBD

B. Results

In this section, the results of the proposed trained network are discussed. The proposed architecture is based on CNN algorithm with two streams, RGB and depth images. Seven possible scenarios for object detection organised as a turtlebot, have been considered. It comes to: "go-straightforward (1)", "turning left with 22.5 degree (2)", "turning- left with 45 degrees (3)", "turning- left with 67.5 degree (4)", "turning -right with 22.5 degree (5)" turning -right with 45 degrees (6)" and turning -right with 67.5 degree (7)". As mentioned above, the classifications of different formats images were performed using two pre-trained CNN algorithms. The images distribution used for training and testing the CNN classifiers is presented in Table II.

TABLE II. THE DISTRIBUTION OF TRAINING AND TESTING IMAGES PER CLASS

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Total
Training	444	360	352	474	348	358	330	2666
Testing	86	71	69	75	67	71	60	499
Total	530	431	421	549	415	429	390	3165

The two-stream CNN architecture use was done in order to evaluate the performance of the classification process of the fusion between RGB and Depth. We generated the confusion matrix associated with RGBD. The confusion matrix is a specific table used to summarise the performance of a classification algorithm. From this matrix, several measures of the performance can be extracted. These include sensitivity, specificity, positive predictive value, negative predictive value and accuracy [15]. The obtained results using a T-CNN architecture are presented in Table III.

TABLE III. CONFUSION MATRIX FOR RGBD

		Target Classes							
		1	**2**	**3**	**4**	**5**	**6**	**7**	
Output Classes	**1**	(82) 16.4%	(0) 0%	(1) 0.2%	(2) 0.4%	(2) 0.1%	(0) 0%	(1) 0.2%	93.1% - 6.9%
	2	(0) 0%	(65) 13%	(2) 0.4%	(0) 0%	(1) 0.2%	(0) 0%	(0) 0%	95.5% - 4.5%
	3	(2) 0.4%	(1) 0.2%	(65) 13%	(1) 0.2%	(2) 0.4%	(1) 0.2%	(0) 0%	91.5% - 8.5%
	4	(0) 0%	(1) 0.2%	(1) 0.2%	(69) 13.8%	(0) 0%	(2) 0.4%	(1) 0.2%	93.2% - 6.8%
	5	(1) 0.2%	(2) 0.4%	(1) 0.2%	(2) 0.4%	(62) 12.4%	(0) 0%	(0) 0%	91.2% - 8.8%
	6	(0) 0%	(2) 0.4%	(0) 0%	(1) 0.2%	(0) 0%	(67) 13.4%	(1) 0.2%	94.3 % - 5.7%
	7	(1) 0.2%	(0) 0%	(0) 0%	(1) 0.2%	(0) 0%	(1) 0.2%	(57) 11.4%	95 % - 5%
		95.1% - 4.9%	91.3% - 8.7%	94.2% - 5.8%	92% - 8%	92.2% - 7.8%	94.3 % - 5.7%	95 % - 5%	**93.4% - 6.6%**

C. Discussion

In this study, we used the toolbox CNN within MATLAB to implement the fusion method. The corresponding confusion matrix is shown in Table III. The accuracy rate for the linear and Gaussian kernels is 93.4%. For better understanding, we consider the diagonal of all tables which show the number of the images correctly classified and their percentages relative to the total number of test images (499). For example, in Table III, 82 RGBD images i.e.16.4% of the total number of RGBD images used for testing, were correctly classified as ‘go-straightforward’ or class 1. Based on table III, this represents 95.1% of the images in class 1 as shown in the first cell of the bottom row.

The off-diagonal cells show the number of misclassified images (and their percentages relative to 499). Using, once again Table 3 as example, cells 2 and 3 of the first row, show that no image (0%) of class 2 were classified as belonging to class 1 and that 1 image of class 3 were classified as belonging to class 1.

The last row represents the percentages of correctly classified images per class. For example, cell 2 indicates that 91.3% (65 out of 71) of the images belonging to class 2 were correctly classified and 8.7% (6 out of 71) of the images of this class were incorrectly classified.

The last column reports the ratios of the “correctly” classified images as belonging to specific classes compared to the “total” number of images assigned by the classifier to these classes. For example, in table III, the number of correct images classified as belonging to class 1 is 82. The total number of images that were assigned to class 1, however is 88 (82+0+1+2+2+0+1). This gives a ratio of 93.1% (82/88). The same reasoning applies to all tables.

A metric that is widely used to when for assessing the performance of classifiers is the accuracy. Accuracy is defined as the total number of correctly classified images over the total number of images. As it can be seen from the confusion matrix, the classifier achieved the best performance with RGBD images with an accuracy of 93.4% (and 6.6% misclassification rate, followed by RGB and depth images).

The proposed model provides better performance in both spatial and temporal flow than the traditional two-stream CNN. Moreover, compared to other artistic approaches, the best overall accuracy can be achieved with the proposed model.

VI. COMPARISON

We compare the performance of the proposed model's two-stream system with others that have already been implemented. Table IV gives a quantitative summary of this comparison results. According to this table, our proposed two-stream model obtains the highest recognition rate among all methods, which is 93.4%. The proposed model shows a performance of more than 8% compared to the two-stream CNN method of [16]. It has also a much better performance compared to the traditional two-stream image classification.

TABLE IV. ACCURACY OF OBSTACLE AVOIDANCE WORKS

Work	Work [18]	Work [19]	Work [17]	Our work
Number of class	5	3	5	7
Number of Inputs	2	1	1	2
Test accuracy	80.2%	81.72%	84.6%	93.4%

VII. CONCLUSION

The paper presented the classification using RGBD (fusion of RGB and Depth) images based on two Streams Convolutional Neural Networks. The essential goal is classifying the images in 7 classes indicating the movement to be taken by the robot to avoid obstacles with a good precision. This work has shown promising results for the fusion method. Overall, the results showed that our system could manage the obstacle avoidance of mobile robots with an accuracy of approximately 93.4% (RGBD) for T-CNN algorithm. Moreover, the proposed system can be used as a platform for further improvements of the accuracy and could manage the obstacle avoidance of mobile robots with an accuracy of approximately 93.4% (RGBD) for T-CNN algorithm. Moreover, the proposed system can be used as a platform for further improvements of the accuracy and adaptability of obstacle detection in diverse environments. Therefore, the CNN trained by our dataset provided classification accuracy in avoiding obstacles in real-time experiments.

ACKNOWLEDGMENT

This research was supported by the project "SAXEED.Fit4Market" funded by the DAAD with funds from the Federal Ministry for Economic Cooperation and Development (BMZ)".

REFERENCES

[1] Ran, Lingyan, et al. "Convolutional Neural Network-Based Robot Navigation Using Uncalibrated Spherical Images." Sensors, vol. 17, no. 6, 2017, pp. 1341-1354.

[2] Lenz, I., H. Lee, and A. Saxena, Deep learning for detecting robotic grasps. The International Journal of Robotics Research, 2015. 34(4-5): p. 705-724Murty, M. Narasimha, and Rashmi Raghava. Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks. Springer, 2016.

[3] S. Berberoglu, “A Review of Image Classification Approaches and Techniques,” Int. J. Recent Trends Eng. Res., vol. 3, no. 3, pp. 1–5, 2017, doi: 10.23883/ijrter.2017.3033.xts7z.

[4] Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review Neural computation." (2017).

[5] L. Tai, S. Li, and M. Liu, “A deep-network solution towards model-less obstacle avoidance,” IEEE Int. Conf. Intell. Robot. Syst., vol. 2016-Novem, no. 6140021318, pp. 2759–2764, 2016, doi: 10.1109/IROS.2016.7759428.

[6] M. Selmi and M. A. El-Yacoubi, “Multimodal sequential modeling and recognition of human activities,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9759, pp. 541–548, 2016, doi: 10.1007/978-3-319-41267-2_76.

[7] C. El Mechal, N. El, A. El, and M. Mesbah, “CNN-based obstacle avoidance using RGB-Depth image fusion,” pp. 867–876, I. Systems, *WITS 2020*. 2020.

[8] S. Zia, B. Yüksel, D. Yüret, and Y. Yemez, “RGB-D Object Recognition Using Deep Convolutional Neural Networks,” Proc. - 2017 IEEE Int. Conf. Comput. Vis. Work. ICCVW 2017, vol. 2018-January, pp. 887–894, 2017, doi: 10.1109/ICCVW.2017.109.

[9] Christian Ertler, H Posseger, M Optiz, and Horst Bischof. Pedestrian detection in rgb-d images from an elevated viewpoint. In 22nd Computer Vision Winter Workshop, 2017.

[10] S. Song, S. P. Lichtenberg, and J. Xiao, “SUN RGB-D: A RGB-D scene understanding benchmark suite,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07-12-June, no. Figure 1, pp. 567–576, 2015, doi: 10.1109/CVPR.2015.7298655.

Educational, Psychological and Behavioral Profiling With High Dimensionality Reduction

Cyryl Koshyk
Technology, AI & Global Expansion (ICT & Innovation Department)
The Kyoto College of Graduate Studies for Informatics
Kyoto, Japan
cyryl@kcg.edu

***Abstract*— In today's era of big data collection and processing, utilizing high end technologies and computing power has revealed many interesting observations and findings embedded in numbers, revealing structure and meaning of that data. Recently, increase in computing capabilities has led to the development of more sophisticated methods for big data processing, solving and prediction.**

In particular, "unsupervised learning" – clustering and high dimensional (PCA [1], Trimap [2], MDS [2]) data reduction new algorithms has emerged (UMAP, T-SNE, CVAE, IVIS) [2], fully utilized and included in Artificial Intelligence workflow.

These technologies and strategies have been employed for numerous big data problems, they are scantly discussed in the context of analyzing survey, and feedback data.

This paper provides an overview of research for educational, psychological and behavioral profiling, specifically focused on manual cluster labeling and high dimensional reduction by using techniques of modern dimension reduction methods, followed by a discussion of how to use the transformed variables in the context of individual profiling solely based on visual input – images in form of perception tests, in attempt to identify, individuals cognitive thinking and reasoning, as well as psychological and emotional profile. Outcome of such provides interesting results, which could be interpreted further to advice and guide individuals for their future choices of – educational subjects, their goals and choosing career. We highlight some of our findings, that were conducted at universities.

Keywords—Education, Artificial Intelligence, Psychology, High Dimensionality Reduction, Profiling, Behavior

I. Introduction

The explosion of big and complex data has resulted in both a dramatic increase in the volume of available data and the possibilities of how to use that data. Artificial intelligence has played significant role in this field, constantly expanding in data processing, filtering, optimization and predicting outputs.

At it is core, data processing is processed to render valuable insight of data meaning and its base structure.

Dimensionality reduction [3] is leading technique, of processing and visualizing complex data, essentially transformation of data from a many attributes or features (high-dimensional space) into low dimension (2d or 3d scatter plots charts in form of clusters) a visual representation so that the low-dimensional representation retains some meaningful properties of the original data.

These clusters shape and form sub-clusters clouds, which sort and gather based on data relationship and meaning of data. Otherwise attempt in analyzing or working in high-dimensional spaces (many attributes and input data) it is near impossible, due many reasons – one of which is computationally intractable. Often dimensionality reduction it is used for noise data reduction, data visualization, cluster analysis which allows for further analysis of data, among it is also used for feature extraction.

II. Data Preparation

From speech, images and sensor input, data from the natural world is often high dimensional. As seen in picture bellow. Digital signal is transformed into raw data, converted using "Feature Extraction", converting audio or speech into meaningful numeric representation, in final dimensionality reduction is being applied, rendering visual clustered scatter plot, allowing to easily identify different sounds or audio (for ex. Spoken numbers) (Fig. 1).

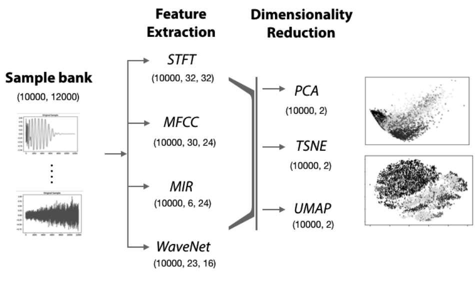


Fig. 1. Data preparation

III. Visual Percepton Test

Our test called "Visual Perception test", was performed, by using only 11 questions, containing 3 to 5 images, it was solely based on visual presentation only along simple question – to approve specific image.

Images were selected specifically, to contain psychological embedding with meaning. Importantly, visuals cause a faster and stronger reaction than words. Which help users engage with the content deeper, and such emotional reactions influence information retention.

Visual memory in humans is encoded in the medial temporal lobe of the brain, the same place where emotions are processed (Fig. 2).

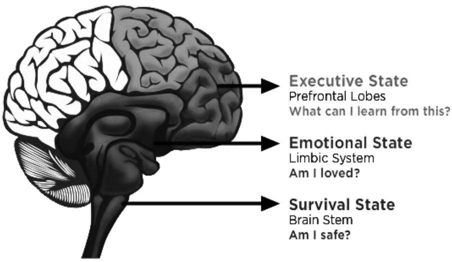


Fig. 2. Visual memory

As example questions related to “Which shape do you like?” (Fig. 3), test subject is presented with 4 images containing different shapes.

Purposely displayed as colorless, black and white, essentially capturing core shape with strong intensity, rendering no influence from colors, which would further distract viewer.

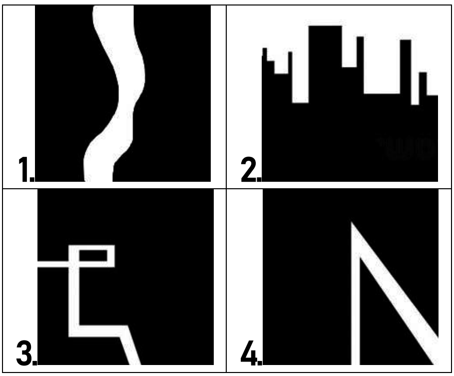


Fig. 3. Example

Essentially shapes, since our childhood, are most basic form of experience to our brains through vision – eyes as receptors. Basic shapes – contain: understanding math and geometric concepts like shape, size, space, and position. Sorting and categorizing, which are essential for problem-solving. Associations with letters and numbers, made of basic shapes. As as visual discrimination to associate shapes and colors with familiar objects.

Following image shapes could be interpreted as following:

- #1 More emotional, sensitive, artistic personality. Non-structured personality, unpredictable cognitive thinking.
- #2 Rigid personality, organized. Sorting, categorizing.
- #3 Distinctive, selective
- #4 Emotional, judgmental, conflictive personality. Students having – tattoos, tendency to violent behavior or enjoying watching cruelty, essentially shape represents – knife, blade, sharp corner object able to hurt.

Images containing shapes from psychological view, contain more embedded features, otherwise impossible to capture just by text only question.

Next sample question named**:** *“What place you would not want to be?”* was based on fear, again only black and white images were presented to individual purposely (Fig. 4).

1. 

2. 

3. 

4. 

Fig. 4. Question - *What place you would not want to be?*

The Fig. 4 -Image #1 would be chosen by person, not be able to solve basic tasks, and finding difficulty in challenges, tree and branches represent complexity itself. Determined person, would appreciate freedom of being “outside” and solving complex problems.

The Fig. 4 -Image #2 would be chosen by person more mature in thinking, person realizing age, image also represents, fear of being in hospital or becoming sick.

The Fig. 4 -Image #3 was chosen by most emotional individuals, it also contains, if you look closely – standing women shape within darkness at bottom of stairs. More sensitive and emotional person, would be able to see it, while less emotional and more leadership personality would not even analyze it that well.

The Fig. 4 -Image #4 was chosen by individual, comfortable in same situations or tasks, it also visually represents “new things” to challenge or overcome, and unknown.

1. 

2. 

3. 

4. 

Fig. 5. Question - *Which are you afraid most?*

Both fear and disgust impart an evolutionary advantage — fear helps us to avoid predators, while disgust steers us away from eating perished plums. These negative emotions are certainly psychological bedfellows, but they're also distinct entities. It is now established that the physiological responses are different: fear activates the sympathetic nervous system, and disgust triggers the parasympathetic nervous system.

Images containing fear, allowed extraction and connection with deeper psychological level with test subject childhood experiences itself, including sparse connection with environment they grew up or what experiences they were exposed to (including – media, TV, environments).

In such data collections, they are more details carried and embedded in visual representation, which creates "high-dimensional data" of individual itself.

By showing to individuals fear based images, it had importance as human emotion that normally, can help protect from danger and prepare to take actions. If people didn't feel fear, they wouldn't be able to protect themselves from legitimate threats. Fear is a vital response to physical and emotional danger that has been pivotal throughout human evolution, as well as shaping human intellectual knowing.

In following question *"Which are you afraid most?"*

Presented images where in color, however two images purposely turned to black and white.

Specifically Fig. 5 -Image #2 – would be chosen by most sensitive person, and Fig. 5 -Image #1 by highest IQ person.

IV. Output Data Preparation

Collected data as answers, where converted and encoded into "one-hot" number. Used in machine learning tasks. As example in picture bellow, animals are being "one-hot" encoded, meaning if answer is present there is number "1", otherwise zero "0" chosen (Fig. 6).

Human-Readable

Pet
Cat
Dog
Turtle
Fish
Cat

Machine-Readable

Cat	Dog	Turtle	Fish
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	0	0	0

Fig. 6. Collected data

For our "Visual Perception Test" image choice where converted to "1".

For example, sample question and answer:

Question 1:	Image #1	Image #2	Image #3
	0	1	0
Question 2:	Image #1	Image #2	Image #3
	1	0	0
Question 3:	Image #1	Image #2	Image #3
	0	0	1

Fig. 7. Sample Q&A

In example above (Fig. 7), OUTPUT data in "one-hot encoded" format would be when merged together, binary combination of:

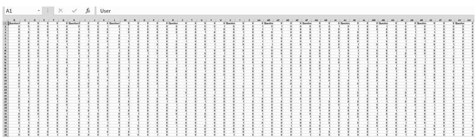
0, 1, 0, 1, 0, 0, 0, 0, 1

In such technique commonly used in machine learning and artificial intelligence, categorical inputs are converted all answer question, as seen in sheet bellow (Fig. 8):

B	C	D	E	F	G	H	I	J	K
Question:1 -	0	0	0	0	0	Question:2 -	0	0	0
0	0	0	0	1	0	0	0	0	1
0	1	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	1	0	0
0	1	0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1
0	0	1	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1
0	0	0	0	0	1	0	1	0	0
0	0	0	0	0	1	1	0	0	0
0	0	1	0	0	0	0	1	0	0
0	0	0	1	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1
1	0	0	0	0	0	0	1	0	0

Fig. 8. Converted data

Combining all questions, as output giving long chain of "0" or "1" number (Fig. 9):



0,0,0,0,1,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,1,0,0,0,0,0,1,0,1,
0,0,1,0,0,0,0,1,0,0,0,0,0,1,0,0,1,0,0

Fig. 9. Chain of 0 and 1

Strain of numbers was generated for every individual participating in "Visual Perception Test".

V. TEST RESULTS

To visualize results, we used recent high-dimensionality reduction technique called – UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) with collected answer data from "Visual Perception Test" we conducted with selected individuals.

UMAP technique it's dimension reduction technique that can be used for visualization, and also for general non-linear dimension reduction. The algorithm is founded on three assumptions about the input data:

- The data is uniformly distributed on Riemannian manifold;
- The Riemannian metric is locally constant (or can be approximated as such);

The manifold is locally connected.

Which gives possibility to model the manifold with a fuzzy topological structure. The embedding is found by searching for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure.

Riemannian geometry to bring real world data a little closer to the underlying assumptions of the topological data analysis algorithm.

Following cluster scatter chart was rendered (Fig. 10).

NOTE: Test subject identification numbers where hidden for privacy reasons

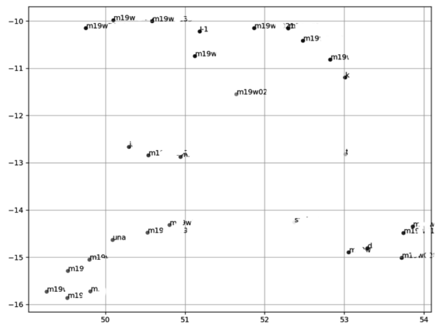


Fig. 10. Cluster scatter chart

After "Visual Perception tests" was completed, each subject was interviewed to confirm cluster chart meaning giving explanatory meaning to sub-groups of clustered groups. Questions where directly asked about their preferences, life style, educational background was confirmed, their goals, fears and education level, including EQ and IQ.

Following final chart was rendered, after MANUAL selection within – latent space (area of possibilities and outcome from input data).

NOTE: Test subject identification numbers where hidden for privacy reasons (Fig. 11)

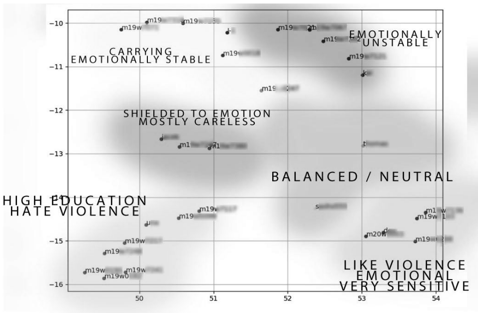


Fig. 11. Test subject identification

VI. CONCLUSION

Output final chart with manual labeling, have presented – emotional and psychological profile of test subjects, as well as their overall educational performance. Interestingly, presented visual images and questions where mostly emotion based containing high EQ [4] engagement, however relatively low IQ, still in our experiment of dimensionality reduction algorithm by utilizing UMAP, precise subject position on the chart was made, placing test subjects in their strongly related similar individuals, ultimately in correct sub-groups

In "green" ("high education" group), all test subjects were against violence, and seeking future career positions as: lawyers, doctors and similar, at other end of cluster map in group "Like Violence, Emotional, Very sensitive" – where individuals not focused in their educational subjects, however – very curious to experience violent image, films or videos and being very judgmental, and jealous by nature in their relationships.

Colored "red" – Shielded to emotion group were individuals also known as "players" type personality, willing to risk and not being afraid. Also being able to change places, or environment without being emotional.

In "violet" group ("emotionally unstable") where group, subjects without leadership skills, shy, unsure, and seeking social approval, that type is closely related to today's famous "Instagram" in Social Network culture, where presentation and social acceptance it is their target.

Assuming by using more questions 50+ personalities will be divided more, giving more insights of test subjects deeper character explanations. If such tests are performed with younger generation, perhaps opens new doors – for career and subject choices, correctly choosing interesting to individual education courses and path.

[1] B.K. Tripathy, “Principal Component Analysis (PCA)”, DOI: 10.1201/9781003190554-2, July 2021.

[2] Christophe Giraudics, “Introduction to High-Dimensional Statistics”, DOI: 10.1201/9781003158745, August 2021.

[3] Shavkat Ayupov, Bakhrom Omirov, “Classification Problem in Low Dimensions”, DOI: 10.1201/9780429344336-3, November 2019.

[4] James Evans, Rebecca Hinds, Malte Friedrich Jung, Niccolo Pescetelli, “Leveraging Artificial Intelligence for Collective Intelligence”, DOI: 10.5465/AMBPP.2021.15722symposium, August 2021

Research on Building Competition-Practice Platform and Promoting the Construction of Electronic-Technology Course

Hui Sun
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
sunhui@hrbust.edu.cn

Haibin Wu*
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
woo@hrbust.edu.cn

Yang Cui
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
4599578@qq.com

Honghao Wang
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
416102049@qq.com

Xiaoming Sun
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
sunxiaoming@hrbust.edu.cn

Dezhong Ma
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
20021741@qq.com

***Abstract*—Electronic technology course includes analog-electronics and digital-electronic technology. It is an important basic course for electronic majors in Colleges and universities. Due to the strong theoretical content of traditional teaching materials and disconnection from specific electronic design practice, electronic technology has always been a course difficult for many students to understand. How to find new methods and new ideas to lead students into a new and interesting field of electronic technology learning has become an important research project in electronic technology teaching reform today. This paper proposes to introduce electronic design competition as a means, build a scientific competition practice platform as a new reform form in electronic technology teaching, maximize students' interest in e-technology learning, and promote students to give full play to their subjective creativity and initiative innovation. It is of great significance to college students understanding and mastering to electronic technology courses under the background of innovation education.**

Keywords—Electronic technology, Practice platform, Teaching reform, Innovation Ability

This work was supported by The Higher Education and Teaching Reform Project of Heilongjiang Province in China (SJGY20190278); The Higher Education and Teaching Reform Project of Heilongjiang Province in China (SJGY20200296); The construction project of high-quality Demonstration Courses for Postgraduates in 2021 of Harbin University of Science and Technology.

I. Introduction

Innovation and Entrepreneurship education is a hot topic in Higher Education in recent years. In many policy documents in China, it is emphasized that innovation and Entrepreneurship education is an important measure to deepen the teaching reform of higher education. Electronic technology course is an important basic course for electronic majors in Colleges and universities [1]. It carries the learning of follow-up courses such as micro-computer principle, digital signal processing and sensor technology. Therefore, electronic technology is also a professional basic course that students in Colleges and universities must be proficient in order to adapt to innovation and entrepreneurship education. At present, due to the strong theoretical nature of the course, and the traditional teaching mode has not changed fundamentally, students understanding and mastery to the course are not ideal [2]. How to explore the new teaching methods of the course and transform the traditional abstract electronic technology theory into professional knowledge that students are willing to learn actively is an important issue to be considered in the teaching reform of electronic technology course under the background of innovation and entrepreneurship education.

II. There are two major problems in electronic technology teaching

A. *The system of traditional knowledge is old and can not meet the requirements of talent training in the new era*

At present, the teaching content of electronic technology course is limited to the traditional old knowledge system, and there is no new content and new theory combined with the times to assist the teaching of the course [3]. The teaching content of the traditional electronic technology course is divorced from modern science and technology. Students lack interest in the teaching content and often be negative to the learning content [4]. Some students have the psychology of curriculum exclusion, no active thinking and innovative consciousness to conduct in-depth research, and even fear of incomprehensible theoretical knowledge. It is an urgent problem to be solved in the construction of electronic technology curriculum to update the curriculum system, increasing the knowledge content integrated with the times, integrating theory into the touchable operation process, so that students can grasp the pulse of the times related to electronic technology [5].

B. *The theoretical content lacks the support of practical links, and can not meet the needs of talent training of innovation and entrepreneurship education*

Most electronic technology courses in Colleges and universities lack practical links supporting theoretical teaching. Students do not understand how to systematically apply the theoretical knowledge of electronic technology in engineering practice [6], how to determine the specific parameters of electronic circuits in engineering projects, and what application details should be considered in project design in electronic engineering. Although some colleges and universities have adopted the project-based learning method, its popularity is limited and can not be applied to extensive promotion. The mastery and application of any theory comes from specific practice. Without practice, there will be no innovation, and the goal of cultivating advanced and professional technology talents in colleges and universities will not be realized [7]. At present, the disconnection between theory and practice has seriously constrained the development of creative thinking of College Students, and it is imperative to build a universal practice platform of electronic technology course [8].

III. Reform and development direction of electronic technology teaching in the new era

A. *Taking the cultivation of professional talents as the guidance, forming the teaching mode of practice assisted teaching*

The purpose of modern education is to cultivate professional talents who meet the requirements of the times [9]. They should not only have a solid theoretical foundation, but also have rich practical ability. Therefore, the teaching content should include both basic theory and practical characteristics of the times. The relationship between traditional teaching content and new knowledge and technology should be coordinated properly, and cutting-edge scientific and technological achievements should be introduced as practical content [10]. Students should participate in the construction of courses and the application of knowledge, and the teaching mode of "teachers teaching" should be changed into the learning mode of "students' doing", so that students can grasp the context of electronic technology courses, and earning interest can be better stimulated.

B. *Building a practical platform for students according to their aptitude, formulating a talent training plan hierarchically, promoting the development of electronic technology courses*

Due to the different qualities of students, their ability to understand and master knowledge is also different. Teaching students according to their aptitude and tapping students learning potentials as much as possible are important goal of modern electronic technology curriculum construction. In order to cultivate students initiative and interest in electronic technology course, it is imperative to formulate a practical skill training scheme of different platforms, according to the knowledge structure and basic abilities of different students [11]. While adhering to the objectives of cultivating high-quality and specialized talents, we should ensure that every student does not fall behind. The practice contents of different majors should be different, and training difficulty should be distinguished according to students' quality. From point to line and from line to plane, establishing a multidimensional comprehensive practice platform, formulating training plan of electronic technology talent hierarchically, which is an important development direction of electronic technology curriculum reform in Colleges and universities

IV. Building a competition practice platform, will improve the learning effect of electronic technology Course Comprehensively

A. *Competition practice platform will promote the realization of the goal of students self-regulated learning in electronic technology course*

To some extent, competition has become the carrier of teaching reform practice and a practical process with special form to realize the goal of high-quality education in colleges and universities [12]. At present, a variety of electronic design competitions have been promoted in science and engineering universities [13]. According to the practical results, electronic competitions have promoted the reform of higher engineering education, especially the reform of practical teaching curriculum system, experimental content and laboratory construction in Colleges and universities.

Take the National Undergraduate Electronic Design Competition and China “Internet Plus” Undergraduate Innovation and Entrepreneurship Competition as examples, students participating in the competition must master the basic theory and related skills of electronic technology to design electronic circuits with certain difficulty [14]. The competition greatly stimulates students initiative in active learning. The mastery to electronic technique course for students has also made a qualitative leap due to the introduction of the competition. Competition practice has become a strong driving force for students self-regulated learning.

B. The competition practice platform will promote the cultivation of creative thinking to students and strengthen the application ability of electronic technology knowledge

While emphasizing the integration of theory with practice, the title of electronic competition needs to meet four combinations: the combination of basic knowledge teaching and comprehensive ability; the combination of calculation during design and test during operation; the combination of basic content and multi-professional content of electronic technology knowledge; the combination of applications of traditional electronic devices and new devices. While emphasizing the basic teaching content of electronic technique, the title of electronic competition also needs to pay special attention to the application of new technology, new devices and new instruments. Electronic competition content should take the application design of analog circuit and digital circuit as the main line, and take into account the overall application of analog-to-digital hybrid circuit, single chip microcomputer, programmable logic device and EDA software tools. Electronic competition process needs to reflect two important connotations: one is to grasp the basic teaching content; the other is to keep pace with the development of science and technology, and constantly provide guidance for electronic technique teaching.

V. A CASE STUDY ON IMPROVING THE TEACHING EFFECT OF ELECTRONIC TECHNOLOGY COURSE BY COMPETITION PRACTICE PLATFORM

At present, after more than ten years of construction, Harbin University of Science and Technology has completed the construction of electronic competition platform and realized the systematization of training. The platform includes multi-level construction consist of national electronic design competition, provincial electronic design competition and school electronic design competition. Practice has proved that the competition platform plays a great role in promoting the knowledge transfer of electronic technology courses. The specific construction contents are as follows.

A. Content construction of competition practice training.

The training contents consists of three processes. The first is the mastery of basic electronic technology knowledge, such as using triode and single power operational amplifier to build amplification circuit. The second is complex circuit design of advanced electronic technology, such as designing multistage active filter to extract small signal without distortion under clutter. The third is the application of CPLD, FPGA and ARM embedded system, such as designing multi-channel analog camera data extraction. There are topics in the training process, such as DC motor speed measurement. There are features, for example, teachers only teaching principles, algorithms and objectives, and students completing the system design scheme, hardware circuit, algorithm selection and program design independently. There is evaluation, such as scoring according to the demonstration effect, and those with good results will be given priority to be rewarded. There are feedback, specific problems and improvement methods encountered in the competition needing to be put forward after the competition.

B. The hierarchical process of competition practice.

Part of the participants of electronic competition are sophomores. Taking Measurement & Control Technology major as an example, all students studying electronic technology courses participate in the training. 187 students participated in the training in 2019, and 220 in 2020. The other part are other grades students who want to acquire more electronic technology knowledge. Still taking Measurement & Control Technology major as an example, 50 students in 2019 and 82 in 2020 participated in the electronic competition. Electronic competition training is organized on the weekends from the fourth week to the fifteenth week of each semester to provide guidance for the production of electronic competition. At present, a continuous tradition and system of electronic competition has been formed. The training process can be attended by all interested students regardless of grade and major. Among them, the participation rate of students majoring in computing and automation is more than 60%. The training method mainly consists of teachers training senior students, and senior students training junior students. In this way, a training mode of “teacher tutor and student tutor” has been formed.

C. Summary of competition practice results.

According to the analysis of the competition results in recent years, the students participating in the electronic design competition have generally achieved systematization and proficiency in mastering the electronic technology knowledge. The electronic competition solved the problem that the electronic technology curriculum is boring and students reject learning, and realized the virtuous cycle of “doing promoting learning, learning to promoting thinking, thinking promoting innovation”. From the examination results of several years, the students participating in the competition practice have a significantly better understanding and answers to the final examination questions of electronic technology than the students not participating in the competition training. Especially for design courses about electronic technology, the project design ideas of the students participating in the competition practice are obviously better than the students not participating in the competition training. From the analysis of feedback information of graduates employment, enterprises give priority to students participating in electronic design competitions with preferential conditions when recruiting graduates, and the employment rate of such students is almost 100%. Electronic design competition makes students recognize the importance of the competition practice to their growth and future work, actively going deep into the ocean of electronic technology knowledge, strengthening learning and practice, consciously improving their engineering practice ability.

VI. CONCLUSION

Electronic technology is an important basic course of professional technology, which involves a large number of basic concepts, theories and circuit analysis methods of electronic science. It is very important for the follow-up courses, including microcomputer principle, digital signal processing and other professional courses. We should study the reform plan of new era of electronic technology teaching, and could build a teaching system to promote the study of electronic technology curriculum through electronic competition practice, which can promote contemporary students to combine the theory with practical application consciously and actively, so that students can better understand the basic theory of electronic technology. It is of far-reaching significance to train students into professional and technical talents in the new era.

REFERENCES

[1] *Mao Q, song Y, Liu D. A preliminary study on the teaching reform of electronic technology course based on College Students smart car competition [J]. Educational observation, 2019,8 (32): 94-96*

[2] *Shangguan J, T. Enlightenment of electronic design competition on the teaching reform of electronic technology [J]. Journal of Yuncheng University, 2019,37 (06): 59-61*

[3] *Tang H B, Guo J X, Niu Quanwang, Zhang Yutao, Chai Yifeng, Liu Mingwei. Promoting digital electronic technology teaching reform through photoelectric competition [J]. Science and technology wind, 2020 (31): 80-81*

[4] *Wu R Y, Yang PQ, Wang Y L, et al.. Curriculum reform of power electronics technology based on electronic design competition [J]. Electronic world, 2021 (02): 77-78*

[5] *Jiang X J, Liu X Y. Practice of project teaching method for the integration of theory and practice of electronic technology [J]. Application of integrated circuits, 2021,38 (09): 66-67*

[6] *Zhu D, Feng C T, Chu K B. Exploration on Project-based Teaching Reform of "electronic technology" curriculum design -- taking digital ultrasonic generator circuit as an example [J]. Education and teaching forum, 2021 (35): 57-60*

[7] *Cui Y, He Zhaolan, Zhang K L, et al.. Practice and exploration of "electronic technology" teaching reform under the background of first-class curriculum [J]. Journal of electrical and electronic teaching, 2021,43 (04): 19-22*

[8] *Ren X L, Wang J. Teaching design of fundamentals of electronic technology under mixed teaching mode [J]. Electronic world, 2021 (15): 41-42*

[9] *Xu H Y, Jiang L, Peng M H, Zhang P Z, et al. Reform of online and offline mixed teaching mode of "electrical and electronic technology" course [J]. Education and teaching forum, 2021 (28): 56-59*

[10] *Hu Q Y, Hu Z J. Exploration on online experimental teaching mode of electronic technology courses [J]. Wireless Internet technology, 2021,18 (13): 138-139*

[11] *Yan D X, Yang J. Research and practice of hybrid teaching mode of "digital electronic technology" course under the background of new engineering [J]. China Arab science and Technology Forum (Chinese and English), 2021 (06): 176-178*

[12] *Dong H L, Fan W. Research on teaching reform of "analog electronic technology" course [J]. Wireless Internet technology, 2021,18 (11): 157-158*

[13] *Zhou D D. Exploration and practice of teaching reform of electronic technology comprehensive practice course [J]. China educational technology equipment, 2021 (10): 91-93*

[14] *Zhang P, Ji Q. Research on the practice and evaluation system of "electronic technology" course under the concept of OBE [J]. Wireless Internet technology, 2021,18 (09): 153-154*

Towards the Perception of Pedestrian Detection Using Vision Based Systems

Julkar Nine
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
julkar.nine@informatik.tu-chemnitz.de

Jayashree Mohan
Automotive Software Engineering
Chemnitz University of Technology
Chemnitz, Germany
jayashree.mohan@s2018.tu-chemnitz.de

Sweta Sharma
Embedded Systems
Chemnitz University of Technology
Chemnitz, Germany
sweta.sharma@s2018.tu-chemnitz.de

Shadi Saleh
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
shadi.saleh@informatik.tu-chemnitz.de

Sanjay Balamurugna
Automotive Software Engineering
Chemnitz University of Technology
Chemnitz, Germany
sanjay.balamurguna@informatik.tu-chemnitz.de

Wolfram Hardt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
wolfram.hardt@informatik.tu-chemnitz.de

***Abstract*— Real-time Pedestrian Detection is one of the critical areas in advanced driver assistance systems (ADAS) and Autonomous Driving. Due to the geometric form of the pedestrian, the variety of the background scene, the moving poses of the pedestrian, and occlusion, pedestrian detection is very challenging. This paper provides a Pedestrian detection algorithm using Image Processing and Machine Learning. The proposed algorithm employs a Support vector machine trained on Histogram of Oriented Gradients (HOG) implemented in a semi-real time environment on an edge device such as Raspberry Pi 3B+. This paper also provides the results and evaluation of our custom dataset with existing datasets, where our custom dataset outperforms existing datasets in terms of accuracy and frame rate.**

Keywords— Raspberry Pi, Pedestrian detection, HOG, SVM, Machine Learning

I. Introduction

Globally, Automobile accidents cause 1.35 million fatalities each year [1]. According to NHTSA, automobile accidents account for 36,096 accidental deaths in the United States as of 2019, where 6205 pedestrians lost their lives [2], and in European Union, 21% of traffic casualties are pedestrians [3]. Owing to the fatalities for the most vulnerable road users like Pedestrians and bikers, initiatives have been undertaken to improve their safety. PROTECTOR ("Preventive Safety for Unprotected Road Users," 2000–2003 [4]) and SAVE-U ("Sensors and System Architecture for Vulnerable Road Users Protection," 2002–2005 [5]) are two such examples of European Commission-funded research initiatives.

Due to this reason, Pedestrian Detection has been a critical research topic and a prime active safety feature of ADAS and autonomous driving. Pedestrian Detection can be efficiently realized through Vision-based Systems. Vision-based object detection for Intelligent Transport Systems gained much attention over the last years due to the extensive research in computer vision, increased processing speeds enabling intensive computational algorithms to run on edge devices, and the increased road vehicles and accidents. Vision-based systems perceive the surrounding environment, such as obstacles, traffic information and, possible collisions through input sensors such as Radar, Lidar, and Cameras. The acquired input data from the sensors are processed and analyzed to enable safe mobility. The widely used sensor in Automotive applications is Radar that detects the obstacle's position and speed but misses to provide the shape of the obstacle. On the other side, LIDAR exhibits high performance in DARPA Challenges in 2005 and 2007 [6], and the obstacle's structure can be extracted from the LIDAR data. LIDAR suffers from few disadvantages like high cost, less reliability during changing weather conditions, etc. In the Automotive Industry, camera-based systems have already moved from development to serial manufacturing, and many advanced driver assistance systems rely on image processing/computer vision [7]. Some of the ADAS features that employ camera sensors are lane departure warning, intelligent headlights control, traffic sign recognition, and our topic of research pedestrian detection, etc. The increased focus on Camera sensors is due to the low cost, minimal power consumption, at least one order of magnitude less than that of lidars, and the 360-degree coverage [6].

Based on visual perception data, the learning, recognition, and classification processes of the Intelligent Transport Systems are carried out through Image Processing, Machine Learning, and Deep Learning. Region of Interest (ROI) Extraction, Feature extraction, and Classification/Prediction constitute the object detection pipeline [8]. This paper aims to deploy the object detection pipeline on Raspberry Pi boards, achieve good accuracy, and simultaneously exhibit real-time processing. We have access to a demonstrator called the CE-box for the target hardware, which was developed under the Computer Science Department of TU Chemnitz [8][9]. The CE box consists of six different slots for sensing and actuating data, and currently, each of the slots is occupied by the Raspberry Pi 3B+ model. Hence, our semi-real-time approach for pedestrian detection focuses on deploying our proposed model on Raspberry Pi 3B+ and further implementation on Raspberry Pi 4B. This objective also concerns the effectiveness of Edge computing, where Self-driving cars are the pinnacle of edge computing. Moreover, it is unreliable to send the car's sensor data to the cloud and receive a response.

Edge computing can be chosen over cloud computing due to real-time response, lower network workload, lower energy consumption, and data security and privacy [10].

The various object detection techniques and their effectiveness are discussed in the next section II. And section III presents the concept of our proposed algorithm, followed by the implementation in section IV. Finally, our proposed algorithm's experiment results and evaluation on different datasets are showcased in section V.

II. LITERATURE REVIEW

The research objective of this paper, pedestrian detection, can also be attributed to object localization as the desired object, i.e., pedestrian, is located on the image. For the task of pedestrian detection, various activities in research can be noticed, which are primarily focused on deep learning and machine learning techniques with various levels of efficiency.

A. *Deep learning*

The process of object detection through Deep Learning is an end-to-end approach, and the feature extraction is automated. The hierarchical feature representations and a reduced number of trainable parameters have made Convolution neural networks more popular among other neural networks in object detection, forming a series of algorithms based on R-CNN. With larger and deeper networks, higher detection accuracy is achieved, proven in several kinds of research. A regression-based method such as the Single-shot detector (SSD) algorithm [11], when combined with Mobilenet architecture, showed a precision and recall rate of 97 percent and 91 percent, respectively, due to the enhancements like multiscale features and depth wise separable convolutions in the model architecture. This paper also demonstrates the effectiveness of a Vision Processing Unit (VPU) for edge device computing. Generally, SSD-Mobilenet architecture is designed to be deployed on other platforms such as embedded and mobile applications [12]. Inspired by this method, S. Nikouei et al. [13] proposed a Lightweight CNN architecture by fusing state-of-the-art SSD architecture with the CNN architecture. This method is proven to utilize less memory than Squeezenet, GoogleNet, and VGG.

Though Deep learning methods achieve high accuracy, they require high computation power, making them less preferable for real-time processing on edge devices. Nevertheless, variations of the YOLO algorithm have overcome this disadvantage. Rachel Huang et al. [14] proposed the YOLO-LITE algorithm, focusing on speed rather than the overall size of the network and parameters. This algorithm provides promising results of running at 21 FPS on non-GPU hardware, which overperforms the SSD-Mobilenet architecture in terms of frame rate. D. Barry. et al. [15] proposed a method for object detection named xYOLO, an adaption of Tiny-YOLO. Reduction in the size of the input layer, the number of filters, and the usage of XNOR on the intermediate layers have made the algorithm run with higher FPS on Raspberry Pi 3B with a slight compromise on the accuracy.

B. *Machine Learning*

Machine learning algorithms consume less energy than deep learning techniques. Separate feature extraction is performed in Machine Learning, followed by selecting a suitable Machine Learning algorithm. Papageorgiou et al. [16] used support vector machines (SVM) to an over-complete dictionary of multiscale Haar wavelets, resulting in one of the early sliding window detectors. The use of gradient-based feature characteristics resulted in considerable benefits. Dalal and Triggs [17], inspired by SIFT [18], popularized histogram of oriented gradient (HOG) features for detection by demonstrating significant improvements over intensity-based features [19]. HOG characteristics have grown in popularity since their debut, with virtually all contemporary detectors incorporating them in some way. Maji et al. [20] developed an approximation to the histogram intersection kernel for use with SVMs, which allows for significant speedups and, therefore, allows for a non-linear SVM in sliding window detection. A texture descriptor based on local binary patterns [21] was coupled with HOG by Wang et al. [22], and a linear SVM classifier was adapted to perform basic occlusion reasoning. There is a boost in performance that is superior to pure HOG. Yunsheng et al. [23] used the HOG feature in combination with the SVM classifier to evaluate the pedestrian cognitive mode in various complicated traffic environments, and they optimized the most effective and efficient machine vision and pattern recognition cognitive technique. He, Miao. et al. [24] extracted the semantic regions of interest by generating the heat map through convolutional neural networks followed by HOG and SVM for classification. This method efficiently detects and classifies the pedestrians with semantic regions of interest and generated results higher than the state of art techniques.

C. *Image Processing*

Image processing forms the initial step of machine learning algorithms for extracting regions of interest (ROI) and the features descriptions. Various image processing techniques like contour detection, clustering, and segmentation actively extract the ROI's that increase the efficiency of the detection algorithm. Having said that, M. Mesmakhosroshahi et al. [25] proposed an ROI generation algorithm for stereo-based pedestrian detection with search space for defining pedestrian candidates. The features like contour density, maximum area, maximum perimeter, and matching score are extracted from the contour information of the color image. These features are then effectively classified by Linear SVM, achieving 98.8% accuracy on the Daimler dataset. Image segmentation is used by Na Shou et al. [26] to extract the ROI's. The proposed method employs Median filtering, histogram equalization followed by a Fuzzy C-means clustering algorithm. And then, a morphological dilation to link the separated regions of the same object is done and classified by SVM. Through this method, the number of detection windows is reduced. On overall analysis of the pedestrian detection algorithms and the research objective on the target hardware, a technique combining image processing and machine learning is presented in this paper.

III. CONCEPT

This paper presents the concept and implementation of the Pedestrian Detection System. The proposed algorithm is a Support Vector Machine trained on HOG features. HOG is well suited for images with more edges. As noticed in the above section II, when combined with SVM, HOG has shown superior performance than other feature descriptors like Haar-classifiers, and it is insensitive to geometric and photometric transformations. SVM is chosen over other machine learning algorithms because it produces significant accuracy while using minimal computing power. The algorithm can be divided into training and prediction stages. Initially, in the training stage, the input images are resized to 64*128 and subsequently converted to grayscale to have reduced computations and more straightforward calculations while splitting the images into 8*8 patches. And then, the HOG features are extracted. Subsequently, SVM is trained on these HOG features. As a result, the SVM model is generated, which is used to detect pedestrians on videos.

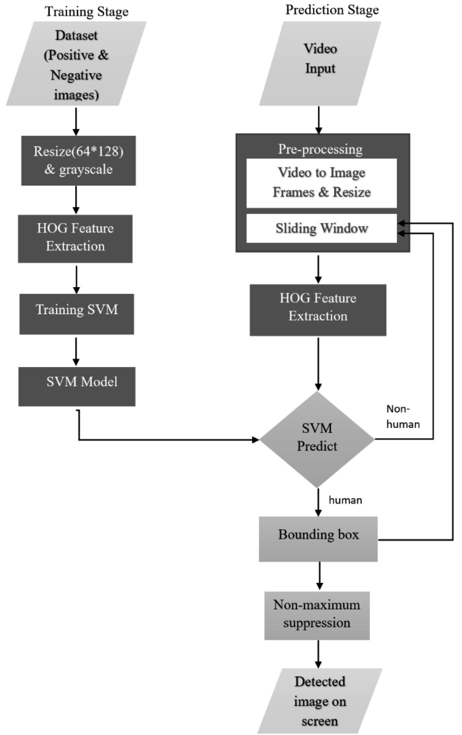


Fig. 1. Flow chart of Pedestrian Detection System

In the prediction stage, the video stream from the camera mounted on the vehicle is divided into image frames and fed into the pre-processing stage. Pre-processing is carried out to convert the original image frames into a more suitable form for the required application. Resizing the image frames is an essential step in pre-processing for pedestrian detection as it is found that the first derivatives of intensity are more significant in downsampled images than in original images, even though the intensity distributions are identical. Following that, a sliding window technique is employed to extract the regions of interest on the complete image frame focusing on the possible regions concentrated with pedestrians. And on each window, HOG feature extraction is applied, and the trained SVM model classifies these HOG features into human or non-human. On detection of a pedestrian, the position of the bounding box is stored, and the algorithm moves to the next image patch of the sliding window. It runs on a loop until the complete image frame is scanned. The direct output from the SVM classifier contains one or more bounding box proposals around a single pedestrian. Therefore, a Non-maximum suppression algorithm eliminates the overlapping bounding boxes with high IOU than the threshold. The flow chart of the proposed algorithm is represented in Fig. 1.

IV. IMPLEMENTATION

The proposed algorithm is implemented on Raspberry Pi 3B+. The operating system used in both is Raspbian Stretch. Several libraries such as OpenCV and Skimage for image processing tasks, Scikit-learn for machine learning operations, and NumPy are installed along with other dependencies. Python is the programming language employed to run the algorithm. The implementation of our algorithm is divided into three stages, and it is shown in Fig. 2.

Fig. 2. Stages of Implementation

A. Dataset

It is essential to have a good dataset selection. Any new feature set must be appropriately verified on suitable datasets to demonstrate its usefulness in the real world. The data sets should be reflective of the applications that are being considered. It is also critical that they do not have any selection biases that might mislead the outcomes. The images that are not accurately labeled and the presence of duplicate images in the datasets are negative information that might affect the training process and make it impossible to attain the desired results. Based on the literature research, the decision to conduct the experiments on the standard datasets such as Daimler and INRIA and a custom dataset to verify the validity of the proposed method was made.

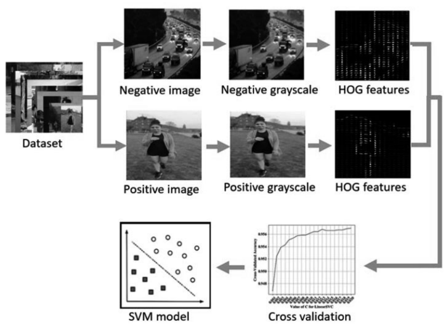


Fig. 3. Data training

B. Data training

The training dataset containing both positive and negative images is pre-processed, where the OpenCV library is used for

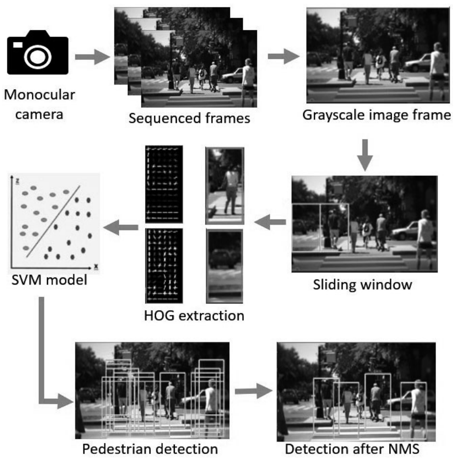


Fig. 4. Classification/Prediction

image resizing and conversion to grayscale. The feature.hog() function from the Skimage library takes in the grayscale image along with the HOG hyperparameters values such as the no. of orientation bins, pixels per cell, cells per block, and the block normalization and generates the HOG features. The extracted HOG features for both positive and negative images are appended into an array with their corresponding labels. Finally, SVM with a linear kernel is trained with the extracted HOG features using the LinearSVC function from the Skicit-learn library on the Google Colab platform. The regularization parameter (C) is chosen using cross-validation. The accuracy of the model for all the C values from 0.01 to 0.2 is assessed. The best value that achieves the highest accuracy is then selected. This training process creates a hyperplane with a maximized margin between human and non-human features, and as a result of data training, we generate our SVM model. The data training stages are shown in Fig. 3.

C. Classification/Prediction

The trained SVM model is imported into Raspberry Pi 3B+. The input video is pre-processed by splitting into image frames, resized using the resize function from the imutils library, and converted to a grayscale image frame. Next, a sliding window function is implemented for extracting the ROI, which takes in the input image frame, the sliding window size, and the step size. As a result, we generate image patches of fixed size 64*128. HOG features are then extracted for each image patch. The predict() function from Python takes in a single argument (HOG feature) and predicts the label based on the trained SVM model. The output contains overlapping bounding boxes around a single pedestrian where each bounding box is associated with the confidence score obtained using decision_function() from the scikit-learn library. The Non-maximum suppression technique from the imutils library is used to eliminate the overlapping bounding boxes. This method inputs the possible bounding boxes with their confidence scores and the overlap threshold. Finally, we obtain a single bounding box around each pedestrian, and the output is visualized on the screen. The data classification/prediction stages are shown in Fig. 4.

V. RESULTS AND EVALUATION

This section presents the experiment results of our Pedestrian detection system on Raspberry Pi 3B+ and evaluation of the system on different datasets along with the performance comparison on different hardware. Our prime focus is on the Custom dataset, a self-made dataset that performed better than a couple of existing datasets.

A. Custom Dataset

The custom dataset is generated after checking the model performance on Daimler and INRIA datasets. Furthermore, it overcomes the drawbacks of Daimler and INRIA datasets in terms of accuracy and frame rate. The custom dataset is created with 3462 images with 2002 positive images and 1640 negative images. The model is evaluated with the performance metrics such as Precision, Recall, and F1 score. On testing our custom dataset model on videos, we achieved an F1 score of around 89% and running in around 5 frames per second on Raspberry Pi 3B+. Fig. 5 shows the example of pedestrian detection using our custom dataset model.

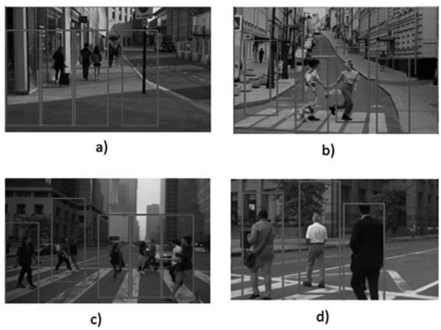


Fig. 5. Examples pf Pedestrian detection using Custom dataset

Cross-validation is performed on the regularization parameter value of the Linear SVM classifier. A range of values from 0.01 and 0.2 are tested, and the best classifier value achieving better accuracy on the validation dataset is chosen, shown in Fig. 6.

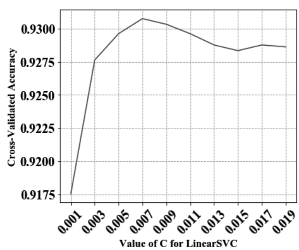


Fig. 6. Cross Validation graph of C

The results from other datasets such as Daimler and INRIA datasets are as follows.

B. Daimler dataset

Daimler dataset contains 24399 images with 14399 positive images and 10000 negative images. The images are very low resolution, making the size very small with respect to the number of images, which makes image processing easier. It suffers a drawback that detection cannot be carried out on videos. Fig. 7 shows the example of pedestrian detection using Daimler dataset.

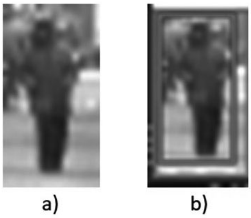


Fig. 7. Example of Pedestrian detection using Daimler dataset

C. INRIA dataset

This dataset contains 2573 images with 902 positive images and 1671 negative images. INRIA model suffers from the issue of Accuracy and FPS trade-off. As the step size of the sliding window greatly influences both the Accuracy and FPS, compromise on FPS is made to maintain the accuracy. Hence, an FPS of around 3 with F1 score of 86% has been achieved on target hardware Raspberry Pi 3B+. Due to low FPS on videos for the INRIA dataset, the Custom dataset is created, which showed better results both in FPS and accuracy. Fig. 8 shows the example of pedestrian detection using the INRIA dataset.

Fig . 8. Example of Pedestrian detection using INRIA dataset

Table 1: Comparison Results

Parameters	Daimler Dataset	Custom Dataset	INRIA Dataset
Penalty Parameters (C)	0.016	0.008	0.1
Kernel	Linear	Linear	Linear
Images / Total Videos	9799	11	8
Total Pedestrians	4799	3368	1822
Detected Pedestrians	4551	2988	1615
False Positive	554	458	287
Miss Rate	8.55%	11.97%	11%
Precision	89.65%	89.86%	84%
Recall	91.44%	88.02%	89%
Accuracy / F1 Score	90.71%	88.93%	86%
FPS on Raspberry Pi 3B	-	4.8	3

The complete evaluation of the three different datasets on Raspberry Pi 3B+ are presented in comparison table 1.

Optimizations such as reduction in step size of the sliding window, employing fixed scaling and reduced sliding window search, and the compromise on resolution can reduce the processing time, increasing the FPS, which is shown through the graph.in Fig. 9 and Fig. 10 presents the variation of FPS on different hardware due to the resolution of the videos on test.

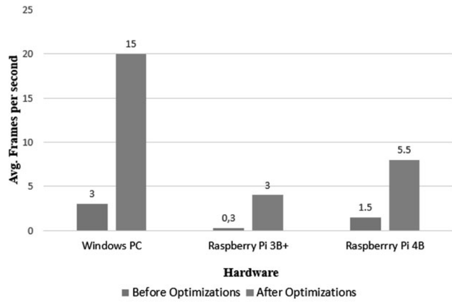


Fig 9. FPS results on Optimizations on different hardware

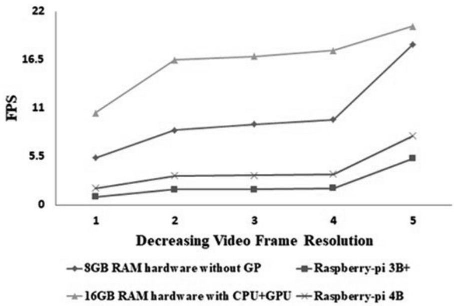


Fig. 10. FPS range on different hardware with decreasing resolution

VI. CONCLUSION

In this paper, an evaluation of the pedestrian detection system by using machine vision is presented. The proposed algorithm was tested on a semi-real-time environment comprising different hardware platforms and three different datasets. All hardware environments have demonstrated proficiency in pedestrian detection but with varying degrees of processing speed. The first contribution shows that combining the HOG feature detector and SVM classifier can be the best algorithm among all other machine learning algorithms to solve pedestrian detection problems. The second contribution of this research paper is that it achieves a high FPS on the Raspberry Pi 3B+ without using a GPU or other additional hardware while maintaining accuracy. A comparison of all datasets by using Precision, Recall, and F1 score was shown in this paper. By evaluating existing datasets with the custom dataset, we were able to identify that our custom dataset performed better in terms of accuracy and speed. To the best of the researchers' knowledge, there has never been a previous study that compared standard datasets with self-made custom datasets. Accurate pedestrian identification may assist drivers in locating pedestrians and reminding them to respond to pedestrians appropriately. The proposed pedestrian detection system can be supported by many Intelligent Transport System (ITS) applications.

In the future, an optimization approach to further increase accuracy and FPS will be used on the proposed algorithm. We also intend to enhance the accuracy of our system even in an unfavorable environment by training and testing our program under adverse weather conditions. Our pedestrian recognition system must be feasible in all situations, including bad weather, excessive light, and nighttime. We also plan to implement the system on NVIDIA Jetson Nano and test it in a real-time environment.

REFERENCES

[1] World Health Organization, "Global status report on road safety 2018. Geneva: World Health Organization; 2018," *Photosynthetica*, vol. 2, no. 1, pp. 1–13, 2018.

[2] "National Highway Traffic Safety Administration," *Federal Register*, vol. 76, no. 237. pp. 76932–76934, 2011, doi: 10.4135/9781544377230.n133.

[3] "Pedestrians | Mobility and transport." [Online]. Available: https://ec.europa.eu/transport/road_safety/users/pedestrians_en. [Accessed: 30-Jul-2021].

[4] V. / Vde *et al.*, "Exploratory Study on the potential socio-economic impact of the introduction of Intelligent Safety Systems in Road Vehicles."

[5] "Save-U project." [Online]. Available: http://www.save-u.org/. [Accessed: 31-Jul-2021].

[6] J. Wei, J. M. Snider, J. Kim, J. M. Dolan, R. Rajkumar, and B. Litkouhi, "Towards a viable autonomous driving research platform," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013, no. Iv, pp. 763–770, doi: 10.1109/IVS.2013.6629559.

[7] G. Ros, S. Ramos, M. Granados, A. Bakhtiary, D. Vazquez, and A. M. Lopez, "Vision-Based Offline-Online Perception Paradigm for Autonomous Driving," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 231–238, doi: 10.1109/WACV.2015.38.

[8] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[9] J. Nine, S. Manoharan, M. Sapkota, S. Saleh, and W. Hardt, "Highway Traffic Classification for the Perception Level of Situation Awareness," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12510 LNCS, no. December, 2020, pp. 286–297.

[10] R. Xu *et al.*, "Real-Time Human Objects Tracking for Smart Surveillance at the Edge," in *2018 IEEE International Conference on Communications (ICC)*, 2018, vol. 2018-May, pp. 1–6, doi: 10.1109/ICC.2018.8422970.

[11] A. Heredia and G. Barros-Gavilanes, "Video processing inside embedded devices using SSD-Mobilenet to count mobility actors," in *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*, 2019, pp. 1–6, doi: 10.1109/ColCACI.2019.8781798.

[12] R. J. Wang, X. Li, and C. X. Ling, "Pelee: A Real-Time Object Detection System on Mobile Devices," Apr. 2018.

[13] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. R. Faughnan, "Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN," in *2018 IEEE International Conference on Edge Computing (EDGE)*, 2018, pp. 125–129, doi: 10.1109/EDGE.2018.00025.

[14] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 2503–2510, doi: 10.1109/BigData.2018.8621865.

[15] D. Barry, M. Shah, M. Keijsers, H. Khan, and B. Hopman, "xYOLO: A Model For Real-Time Object Detection In Humanoid Soccer On Low-End Hardware," in *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2019, vol. 2019-Decem, pp. 1–6, doi: 10.1109/IVCNZ48456.2019.8960963.

[16] [C. Papageorgiou and T. Poggio, "Trainable system for object detection," *Int. J. Comput. Vis.*, vol. 38, no. 1, pp. 15–33, 2000, doi: 10.1023/A:1008162616689.

[17] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[18] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[19] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012, doi: 10.1109/TPAMI.2011.155.

[20] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587630.

[21] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[22] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, no. Iccv, pp. 32–39, doi: 10.1109/ICCV.2009.5459207.

[23] L. Yunsheng, C. Jie, C. Xuewen, Z. Feng, and L. Jingling, "Auto-recognition Pedestrians Research Based on HOG Feature and SVM Classifier for Vehicle Images," in *2020 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2020, pp. 304–309, doi: 10.1109/RCAR49640.2020.9303268.

[24] M. He, H. Luo, Z. Chang, and B. Hui, "Pedestrian Detection with Semantic Regions of Interest," *Sensors*, vol. 17, no. 11, p. 2699, Nov. 2017, doi: 10.3390/s17112699.

[25] "FEATURE-BASED ROI GENERATION FOR STEREO-BASED PEDESTRIAN DETECTION Maral Mesmakhosroshahi , Maziar Loghman , Joohee Kim Illinois Institute of Technology Dept . of Electrical and Computer Engineering , Illinois , USA," *IEEE Int. Conf. Acoust. Speech, Signal Process. 2017*, pp. 1727–1731, 2017.

[26] N. Shou, H. Peng, H. Wang, L.-M. Meng, and K.-L. Du, "An ROIs based pedestrian detection system for single images," in *2012 5th International Congress on Image and Signal Processing*, 2012, pp. 1205–1208, doi: 10.1109/CISP.2012.6469927.

Feature Points Detection and Matching in Lumen 3D Reconstruction Based on SLAM

Kaiyang Xu
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology
Harbin, China
920610061@stu.hrbust.edu.cn

Haibin Wu*
Heilongjiang Province Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation, Harbin University of Science and Technology
Harbin, China
woo@hrbust.edu.cn

Ying Wang
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology
Harbin, China
wangwang13222@163.com

Jianbo Zhao
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology
Harbin, China
740414505@qq.com

Yan Zhang
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology
Harbin, China
1303099014@qq.com

Ruotong Xu
Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology
Harbin, China
xrt8023@163.com

***Abstract*—In this paper, the detection and matching method of feature points in lumen of SLAM (Simultaneous Localization and Mapping). Firstly, the lumen image is preprocessed to eliminate lumen moisture and mirror reflection. Secondly, an improved FAST (Features from Accelerated Segment Test) algorithm was proposed to detect vascular branch points, which can avoid detecting residual highlights and increase point cloud density. At the same time, the branch segments of blood vessels were detected by recursive tracking. On the basis of block matching, ZSSD (Zero Sum of Squared Differences) is used for feature points matching, and line matching based on vascular structure is done. Compared with the classical algorithm, the repetition rate increased by 5.67%. Finally, Hamlyn and Pelvis data sets were used for 3D (Three Dimensional) reconstruction of the human cavity, and the results of 3D reconstruction were quantitatively evaluated.**

Keywords—SLAM, feature points detection, blood vessel bifurcation point, 3D reconstruction

I. INTRODUCTION

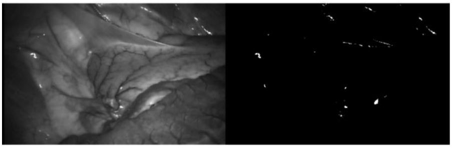
With the rapid development of modern medical technology, medical imaging technology is widely used in the field of clinical medicine, which provides an important basis and indispensable means for doctors' clinical diagnosis and treatment evaluation [1]. However, due to the limitation of 2D(two-dimensional) imaging display mode, the location of surgical target and surgical approach can only be determined by combining surgical experience and the location of lesions in memory during surgery, which greatly increases the difficulty of surgery and is not conducive to the safety and convenience of surgery [2]. The current body surface projection method of patients projects 2D lumen images. It can reflect the inner cavity structure and spatial relationship more accurately and truly, if a 3D lumen model can be established and projected [3]. At present, 3D modeling technology is mature, but there are many limitations for rapid 3D modeling in narrow and characterless inner space [4-5]. In this paper, SLAM is used to realize 3D reconstruction of lumen based on endoscope video sequence. However, moisture and mirror reflection in the lumen seriously interfere with image feature extraction and registration, so a novel lumen feature system needs to be constructed. In this paper, the unique characteristics of the lumen image were considered, the lumen image preprocessing, feature points detection and matching are implemented to complete 3D reconstruction model of the lumen.

II. IMAGE PREPROCESSING

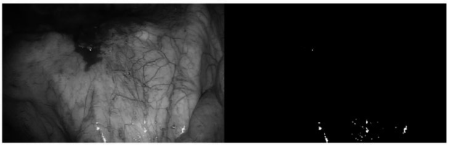
The wet and smooth environment of the cavity in human body results in a great difference between the 3D reconstruction of the cavity image by SLAM and that of the indoor image. Before feature points detection, it is necessary to preprocess the lumen image of human body. In this paper, lumen images were selected from Hamlyn and Pelvis medical datasets respectively,and the main preprocessing content is specular suppression and vascular morphological enhancement of lumen image.

A. Specular suppression

Based on the automatic segmentation and repair method of specular highlights in endoscope imaging proposed by Saint-Pierre [6], the detection is divided into absolute bright area detection and relatively bright area detection by taking full account of the color and gradient changes of the specular area. The absolute speckle region is detected by the color characteristic of the speckle pixel, and the relative speckle region of the inner cavity image is detected by median filtering, as shown in Fig. 1.



a) Original Hamlyn data set with its highlight detection



b) Original Pelvisdata set with its highlight detection

Fig. 1. Highlight detection

This paper improved the FMM(Fast Marching Method algorithm) [7], The image is modified by replacing all the detected highlights with the centroid color of the pixels within a certain distance of the contour. Gaussian kernel function (σ =8) is used to filter the modified image to avoid excessive smoothness of the original algorithm, and a highly smoothness image without highlights is obtained. Then the binary mask of the highlighted region in the image is marked and transformed into a smooth-weighted mask by adding a nonlinear attenuation to the contour of the highlighted region. As shown in Fig. 2, it can suppress most of the highlights and maintain the clear texture and outline of the image.

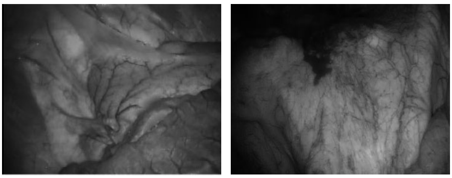


Fig. 2. Specular repair results

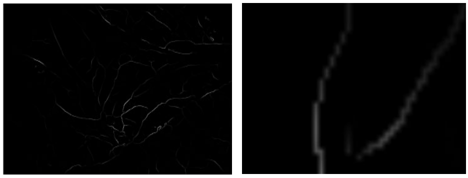
As shown in Table I, the number of specular pixels before and after specular repair was counted. The number of specular pixels decreased from 5.76% to 0.11% and from 2.84% to 0.07%, respectively. Through the ratio of specular pixels, the suppression effect of specular restoration on specular was illustrated.

TABLE I. HIGHLIGHT REPAIR

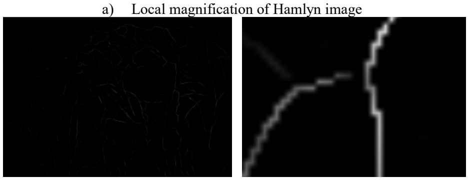
Image	Specular repair	Number of high-resolution pixels	Total number of pixels	Proportion decline rate (%)
Hamlyn Image	no	17701	307200	5.76
	yes	343		0.11
Pelvis Image	no	26118	920000	2.84
	yes	647		0.07

B. *Vascular enhancement*

n RGB fundus images, the contrast between the blood vessels and the background is the best and the blood vessels are the clearest by using green channel for image graying [8]. The enhanced filtering algorithm based on Hessian matrix is used to distinguish tubular structures.



a) Local magnification of Hamlyn image



b) Local magnification of Pelvis image

Fig. 3. Blood vessel monopixelation

III. FEATURE POINTS DETECTION AND MATCHING

In order to evaluate camera motion using images for 3D reconstruction. Feature points need to be detected first, so that the same corresponding points can be found in different frames, even if the camera Angle of view has a small amount of rotation or displacement in different frames of the video stream. Lumen features are single and repeatable, and the extraction and matching of feature points need to consider the special environment of lumen, which needs to be adjusted and improved on the basis of classical algorithm. To establish a special detection scheme for the lumen image of vascular morphology in this paper.

A. *Feature Points Detection*

The point cloud formed by the feature points obtained by CDVB(Circle Detection of Vascular Branches algorithm) only depends on the number of vascular branch points. When the branch points are sparse, the accuracy and effect of subsequent 3D reconstruction will be affected. The similarity between the principle of FAST algorithm and CDVB algorithm lies in that the pixel value on the circle should be judged after the circle is made. However, because the CDVB algorithm is to find independent peak values on the circumference pixel, and the FAST algorithm limits the continuous pixel values on the circumference. The repetition rate of the feature points detected by FAST algorithm and CDVB algorithm is zero, which is mutually exclusive in the detection results. Because of the high computational efficiency of FAST, this method is selected to supplement points.

In the FAST algorithm, if a certain number of pixels around a pixel are different from the pixel value of the point, the point is considered to be a corner point [9]. As shown in Fig. 4, point P is a candidate detection target point and its pixel value is I_p, and r is the radius of the circle. Set a threshold value $\mathcal{E}_d$, the difference between the 16 pixel values of the circle and the pixel values of point P is used to obtain $I_p - I_x$, and compare this value with the threshold value $\mathcal{E}_d$, as shown in (1). there are N consecutive pixels out of 16 that are more different from I_p than $\mathcal{E}_d$, then the candidate point will be judged as a feature point. The larger $\mathcal{E}_d$ is, the fewer points meet the condition.

$$N = \sum_{x\forall(circle(p))} \left| I_x - I_p \right| > \varepsilon_d \tag{1}$$

$$S_x = \begin{cases} \text{darker}, I_x \le I_p - t \\ \text{similar}, I_p - t < I_x \le I_p + t \\ \text{brighter}, I_p + t \le I_x \end{cases} \tag{2}$$

Where t is the threshold value that according to different image environment, the default value is usually 20. I_p is the pixel value of the center, and I_x is the pixel value of points on the circle. In this way, the pixel value on the circumference is divided into three cases, take the value of S_x. When the number of dark spots or bright spots continuously reaches greater than N, then the point is considered as a candidate corner point. The common value of N is 9 or 12, as shown in Fig. 4.

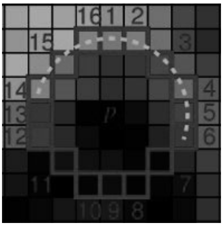


Fig. 4. The image location

However, this algorithm has the problem about uneven distribution of feature points, especially the accumulation of local feature points. There are two reasons for this kind of accumulation: first, there are multiple effective corner points in a small area; second, due to the width of blood vessels, it is likely to detect multiple feature points in an effective corner area, except for one effective feature point, the others are redundant. As shown in Fig. 5, the redundancy of two feature points is marked with a red rectangular box. This situation not only increases the computational burden, but also reduces the risk of feature points matching accuracy. In the above case, consider applying the NMS(Non-maximum suppression) algorithm to eliminate redundancy of feature points.

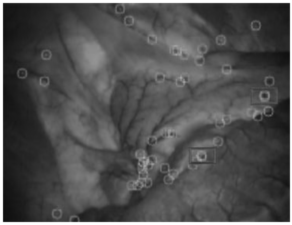


Fig. 5. Feature points redundancy

As shown in Fig. 6, due to incomplete highlight suppression, After the threshold value of FAST algorithm is reduced, specular pixels are detected. Such points vary with the orientation and Angle of camera and light, and cannot be used as effective feature points to participate in subsequent matching and 3D reconstruction. Due to the particularity of the lumen environment in human body, its texture features are blood vessels, which are dark after gray scale, other backgrounds are light. The feature points are attached to blood vessels, so the feature points are dark and the gray value is small. The highlight area is large compared with the pixel values in the pixel neighborhood. Therefore, the FAST algorithm is improved to eliminate the highlights which are mistaken as feature points, according to the lumen environmentin

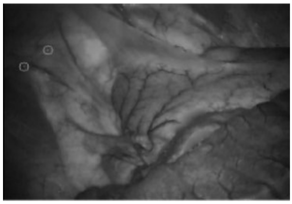
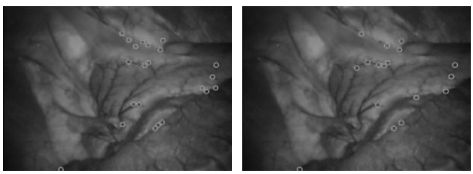


Fig. 6. Residual highlight area detected by FAST

When comparing I_p and surrounding pixel values, only the case $I_x - I_p > t$ is considered, as shown in (3).

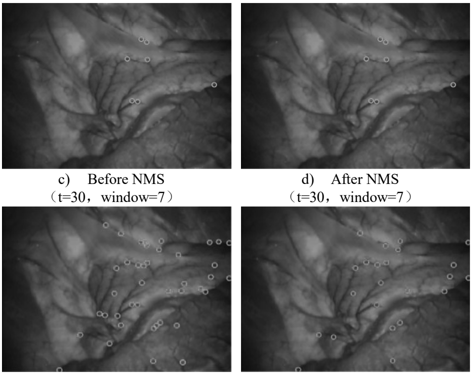
$$N = \sum_{x\forall(circle(p))} (I_x - I_p) > \varepsilon_d \tag{3}$$

The improved FAST algorithm was used to test in the Hamlyn image, and the results are shown in Fig. 7.



a) Before NMS (t=25, window=7)

b) After NMS (t=25, window=7)



c) Before NMS（t=30，window=7） d) After NMS（t=30，window=7）

e) Before NMS（t=20，window=5） f) After NMS（t=20，window=5）

Fig. 7. Improved fast feature points detect results

According to the above figure, it can be seen intuitively that the feature points marked by circles do not contain specular pixels, which successfully eliminates the influence of incomplete specular suppression.

B. Branch Detection

The branch segments of vena cava were divided into full branch segment and half branch segment. The full branch segment is the vascular segment between two fulcrum, and the half branch segment is the vascular segment between the branch point and the vessel endpoint. Both types can be used as a starting point for vascular tracing to achieve detection purposes. As shown in Fig. 8, the green line is full branch segment, the blue line is half branch segment.

Fig. 8. Detection of lumen branches

C. Feature Matching

Feature matching is a key link in SLAM framework.

1) The block matching method adopted is based on the feature points neighborhood block matching algorithm in PTAM(Parallel Tracking And Mapping). The sum of squares of error is calculated in the neighborhood of matching points, and the matching point pairs are determined by the sum of squares of error. This article is based on the block search process, first saving information from previous frames by defining a search area for each map point in the current frame. For each point P that has been saved in the frame, its corresponding point Q in the current frame can be obtained by homography mapping. The search area of P is a circular area centered on Q, and its radius is 1/20 of the image width. Secondly, the feature points located in the neighborhood in the current frame are selected, which are called neighboring points. The neighborhood block of 21×21 size at each neighboring point is compared with the neighborhood block of equal size at point P.

Affine warping obtained from ground truth homography maps is applied to local blocks of P to account for viewpoint changes and the ZSSD of each block is calculated. Finally, if a neighboring point has a minimum ZSSD value which is less than the predefined threshold of 0.02, it is considered a match. The condition for determining the correct pair of corresponding points is $|Q' - Q| < 3.5\text{pixel}$. In Fig. 9, 44 point pairs connected by blue lines are correct corresponding point pairs, and 9 point pairs connected by red lines are wrong point pairs. The matching accuracy rate (repetition rate) was used to illustrate the merits of the matching method, which was obtained by dividing the correct number of matches by the total number of matches, and the matching rate was 83.02% under the data set. The high matching accuracy indicates that the feature points detector has high repeatability and the extracted image blocks have good matching effect.

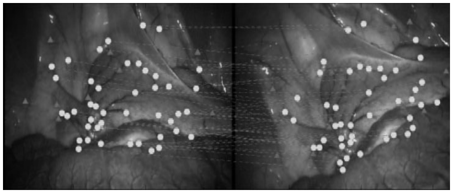


Fig. 9. ZSSD feature points matching

2) According to the texture characteristics of lumen image, not only branch points but also branch segments of vascular segment are detected. In the matching, the special geometric structure of vascular lines can be used to assist the block matching based on feature points. The matching relation between pixels on the branch segment can be established by matching the branch segment of blood vessels. Take two frames for matching processing, on the premise that the branch point matching is completed, the pixel M on the branch segment in key frame I can obtain the corresponding matching point, according to the intersection point of the corresponding polar line and the right branch segment. After detecting the branch segment of blood vessel, for the pixel M, the corresponding pixel M' on the branch segment in key frame II was detected. In this method, the parallax between the matching points is used to search the matching points on the branch segment, so that the matching points need not be checked one by one along the pole line [10].

Since the parallax of the obtained branch points is an estimated value, there are also deviations in pixel coordinates of the matching points on the corresponding branch segment. Note that this deviation will lead to the overall deviation of the corresponding polar line, so the second search needs to correct the actual position of the polar line, so that the pixel matching on the branch segment starting from this branch point can be uniformly corrected. The obtained predicted matching point is set as point F, and a ± 3 pixel wide polar line band is set. The accurate polar line position is determined through a secondary search within this range, as shown in Fig. 10, so as to determine the accurate matching point position.

Fig. 10. Point matching correction of vascular branches

The blue dots are the predicted matching points, the red dots are the exact matching points obtained after the second search.

In this paper, video sequences from the Hamlyn medical dataset were used to conduct comparative experiments of feature point detection and matching.

SIFT (Scale-invariant feature transform) mainly uses DoG(Difference of Gaussian) to realize scale invariance of feature points and local gradient value of image to realize rotation invariance. SURF (Speeded Up Robust Features) is an accelerated version of SIFT algorithm proposed by Hebert Bay in eccV in 2006, and the matching of objects in two images is basically realized in real time under moderate conditions [11]. SURF algorithm ensures the same precision of feature points as SIFT algorithm, and achieves several times higher efficiency than SIFT algorithm. ORB (Oriented Fast and Rotated Brief) is based on FAST algorithm and introduces direction calculation method based on BRIEF (Binary Robust Independent Elementary Features) feature descriptor [12].

Table II makes statistics of the number of feature points detection and matching repetition rate under SIFT, SURF, ORB and the algorithm in this paper. These two indicators are used to describe the characteristics of each algorithm.

In different scenarios, different algorithms have their own advantages according to the differences of environmental characteristics. For example, SIFT algorithm shows excellent performance when it requires high accuracy but no real-time performance. SURF improves algorithm efficiency on the basis of guaranteed performance, while ORB algorithm's advantages are mainly reflected in algorithm efficiency. The algorithm in this paper is second only to ORB in the computation time while maintaining the repetition rate.

In the matching of lumen images, vascular features are superior to general features, and the matching repetition rate further verifies the significance of vascular branch point features. In block matching, the pattern of branch points is more unique than that of vessel points. Therefore, through quantitative comparison, it is concluded that the detection and matching based on vascular features make the repeatability of feature points between different frames higher than SIFT and SURF algorithms.

TABLE II. COMPARISON OF FEATURE POINTS DETECTION AND MATCHING ALGORITHMS

Text	**Algorithm**	**Feature points detection and matching**			
		Fig A. Number of feature points	***Fig B. Number of feature points***	***Repeat rate***	***In summary(ms)***
Group A	SIFT	85	109	85.72%	1312
	SURF	51	46	82.55%	898
	ORB	69	71	81.31%	382
	CDVB	53	59	83.02%	554
Group B	SIFT	179	193	83.21%	1556
	SURF	142	128	81.55%	946
	ORB	161	180	79.57%	562
	CDVB	137	140	87.93%	682

IV. 3D RECONSTRUCTION

SLAM was proposed by Durrant-Whyte, and it is mainly used to solve the problems of localization, navigation and map construction of mobile robots running in unknown environments [13]. Based on the particularity of the branch points of the cavity image studied in this paper, the feature points method of SLAM visual odometer is used to estimate the camera pose. Camera pose estimation is a problem to estimate the position and attitude of the camera shooting the frame according to the selected key frame.

After the feature point matching above, several matched point pairs can be obtained, and the coordinate data of each set of point pairs can be obtained. Based on the coordinate values of two matching points, a 2D-2D algorithm is used to estimate camera pose in the background of monocular SLAM. 2D-2D indicates that the key frame under a single item is a two-dimensional image, so only the two-dimensional coordinate information of feature points can be obtained. For a pair of matching points there are two two-dimensional coordinates. Since this paper is carried out under the condition of monocular camera, every frame of video stream is 2D image, and the feature point at this time is the point in 2D coordinate, and 2D-2D is used for matching. During the initialization of monocular SLAM, the known condition in 2D-2D case is the two-dimensional coordinates of the point clouds of the two key frames.

After feature points extraction, point clouds formed by feature points will be obtained for the image of lumen in each key frame. Point clouds from different perspectives will be fused by camera pose to form a larger point cloud. After the point cloud of the lumen image is obtained, the points are connected to the surface by Poisson surface reconstruction algorithm. This Poisson based surface reconstruction method has many advantages. Many implicit surface fitting methods segment the data into several regions for local fitting, and further combine these local approximations using mixing functions. But Poisson algorithm is based on the global, the complete data is taken into account all at once. as shown in Fig. 11.

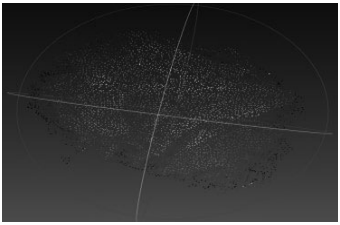


Fig. 11. Dense point clouds ues Hamlyn data set

Complete data processing can avoid the jump of junction caused by local reconstruction and make the reconstructed surface smooth and smooth. However, because the ideal RBF network has global support and non-attenuation, Poisson method allows a local support function hierarchy, so its solution can be reduced to a well-conditioned sparse linear system, as shown in Fig. 12. And then the RGB image is fragmented and affixed to the surface for texture mapping, so as to complete the 3D reconstruction of lumen, as shown in Fig. 13.

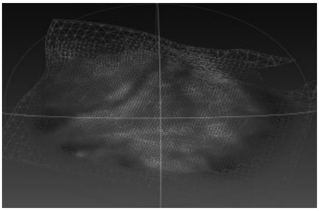
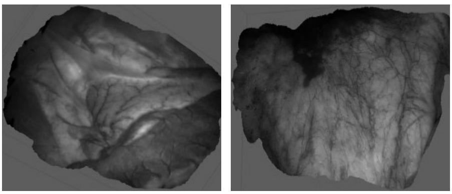


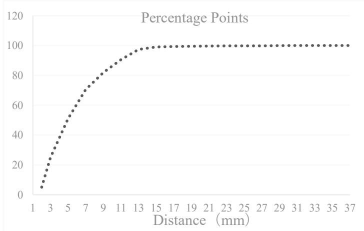
Fig. 12. Surface reconstruction



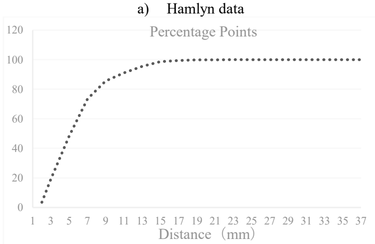
a) Hamlyn data b) Pelvis data

Fig. 13. 3D reconstruction of lumen

In this paper, the endoscope datasets of Hamlyn and Pelvis were used. The two video sequences selected were of different complexity and varied in vascular density. Six key frames were selected for two image sequences respectively. RMSE (Root Mean Square Error) of 2.20mm and reconstruction rate of 71.59% were obtained for Hamlyn sequence and RMSE of 1.94mm and reconstruction rate of 74.63% for Pelvis sequence, Fig. 14 is euclidean distance distribution between surfaces.



a) Hamlyn data



b) Pelvis data

Fig. 14. Distance distribution

V. Conclusion

In this paper, the detection and matching method of human lumen feature points under the framework of SLAM is studied, and finally complete 3D reconstruction of lumen by Hamlyn and Pelvis medical data sets.

For the image preprocessing of the lumen, the specular reflection caused by the wet lumen environment is solved through the specular restoration. Hessian image enhancement was performed for the morphology of blood vessels, and single-pixelated blood vessels were processed according to first-order difference. For feature points detection and matching of lumen images, the improved FAST algorithm was used to eliminate the specular pixels to make the point cloud more compact. Branch points are used to track branch segments recursively. On the basis of block matching, neighborhood block matching of ZSSD is carried out, and line matching based on vascular structure is studied and its repetition rate is improved by 5.67% compared with the classical algorithm. Finally, the key frames in the video sequence of the data set were intercepted for experiment, and the 3D reconstruction of the lumen was completed. Using the reconstructed surface of CT scan as the standard, RMSE of Hamlyn sequence is 2.20mm and the rate of reconstruction is 71.59%. RMSE of Pelvis sequence is 1.94mm and the rate of reconstruction is 74.63%, which proves the feasibility of the proposed method. At the same time, it lays the foundation for the subsequent projection of the body surface projection method to display the 3D surgical scene.

References

[1] H. B. Wu, R. T. Xu, and A. L. Wang, “Overview of 3D human lumen reconstruction technology based on computer vision,” Computer Engineering, vol. 47(10), pp. 1-16, 2021. (in Chinese)

[2] T. M. Wang, X. H. Zhang, and X. B. Zhang, “Review of research progress in laparoscopic augmented reality navigation,” Robot, vol. 41(01), pp. 124-136, 2019. (in Chinese)

[3] P. Liu, P. F. Shao, and J. M. Ma, “A co-axial projection surgical navigation system for breast can- cer sentinel lymph node mapping : system design and clinical trial,” SPIE BiOS. Proc SPIE 10868, Advanced Biomedical and Clinical Diag-nostic and Surgical Guidance Systems XVII, San Francisco, California, USA, 1086, pp. 48-53, 2019.

[4] J. Lamarca, S. Parashar, and A. Bartoli, “DefSLAM: Tracking and Mapping of Deforming Scenes from Monocular Sequences,” IEEE Transactions on Robotics, vol. (99), pp. 1-13, 2020.

[5] Christophe, Doignon, and Jose, “Live Tracking and Dense Reconstruction for Handheld Monocular Endoscopy,” IEEE Transactions on Medical Imaging, vol. 38(1), pp. 79-89, 2019.

[6] C. A. Saint-Pierre, J. Boisvert, and G. Grimard, “Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images,” Machine Vision & Applications, vol. 22(1), pp. 171-180, 2011.

[7] T. Yuan, X. L. Zhou, and X. F. Wang, “Registration and occlusion handling based on the FAST ICP-ORB method for augmented reality systems,” Multimedia Tools and Applications, 2021 (prepublish)

[8] G. Ranjit, S. Aditya, and D. Sayan, “An improved vessel extraction scheme from retinal fundus images,” Multimedia Tools and Applications, vol. 78(18), 2019

[9] Y. Liao, W. H. Huang, and H. Zhong, “Tunnel Layout Map Stitching Method Research Based on FAST Feature Detection and FLANN Feature Point Matching Algorithm,” Transportation energy saving and environmental protection, vol. 17(02), pp. 137-141, 2021.

[10] N. Mahmoud, T. Collins, “Live Tracking and Dense Reconstruction for Handheld Monocular Endoscopy,” IEEE Transactions on Medical Imaging, vol. 38(1), pp. 79-89, 2019.

[11] H. Bay, T. Tuytelaars, and L.V. Gool, “SURF: Speeded up robust features,” Proceedings of the 9th European conference on Computer Vision - Volume Part I. Springer-Verlag, 2006.

[12] Y. Chen, Q.H. Zeng, and Y.S. Wang, “An Improved ORB Algorithm Based on Edge Feature,” Navigation and control, vol. 17(2), pp. 18-24, 2018. (in Chinese)

[13] J.J. Leonard, H.F. Durrant-Whyte, “Mobile robot localization by tracking geometric beacons,” IEEE Transactions on Robotics & Automation, vol. 7(3), pp. 376-382, 2002.

Situation Analysis of ICT Incorporation Into Education and Teaching

Lkhagvasuren Purev
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
p_lkhagva@mier.mn

Buyandelger Banzragch
Mongolian institute for Educational research
Ulaanbaatar, Mongolia
buyandelger@mier.mn

Nyamsuren Boloo
Higher Education and TVET Research Sector, MIER
Ulaanbaatar, Mongolia
nyamsuren@mier.mn

Abstract: **In line with globalization and the information highway, the Mongolian education system has been developed rapidly to introduce information systems and ICT equipment and tools to the education sector since 2000. Nowadays this pandemic circumstance, the Mongolian education sector is managing issues in a short period of time, however, according to the situation analysis, there is a lot of consideration to be taken in the future. The aim of this study is the based on current state to identify needs for further concern to improve the ICT equal access to quality education. In determining the effectiveness of ICT in education projects and programs, it is worth assessing tangible outcomes rather than simply looking at surface factors such as the number of computers and internet speed. There is also a need to research ICT need in Non-Formal Education in order establish the conditions for quality, equitable lifelong education.**

Keywords: Quality, equal access to education, education policy, ICT, training, internet, Higher Education Institution, competency.

INTRODUCTION

Over the last 20 years, the Mongolian government has placed significant attention to the incorporation of ICT into education, implementing a number of comprehensive projects and programs in that pursuit. Having closely examined the 26 policy documents and papers on the regulations, programs and projects involving the incorporation of ICT into education implemented since 2000, this paper divides the process into the following 4 phases:

1. *2000-2005: Introducing ICT equipment and tools to the education sector.*

The goals and activities encompassed in the policy documents in this phase entail the development of distance education, introduction of ICT competency to civil servants, secondary school instructors and staff, establishment of training centers, increasing the number of computers as well as TV's and other equipment in secondary schools with a special emphasis placed on providing internet connection to educational institutions.

2. 2005 - 2010: Establishing an ICT infarstructure in education activities.

This phase involved further pursuing the previous objectives in addition to introducing an Education Sector Information System (ESIS) into elementary and secondary education, establishing an information database and an ICT infrastructure, implementing a national innovation framework and developing advanced technologies through the establishment of an ICT human resource base. The ICT incorporation policy and objectives involved advanced professional development opportunities for instructors, providing laptops to every instructor, and establishing a learning environment with ICT equipment.

3. 2010 - 2015: Establishing the conditions for open ICT and education activities at the national and international level.

The policy documents and papers in this phase established a requirement for instructor preparation programs to equip new instructors with basic ICT competencies and encompassed the creation of open and distance training centers to offer programs at varying levels of competence. Building on this, the development of online training based on the needs of citizens as well as programs to promote the ethical and safe use of ICT were also encompassed within the objectives in this phase. This points to a commitment to increase the use of ICT and enhancing the competence of users. On the other hand, this phase also saw an emphasis being placed on the integration of ICT into education, establishment of an ICT base for the professional development of instructors as well as supporting knowledge exchange and cooperation in ICT.

4. 2015 - present: Establishing and developing a comprehensive set of services based on ICT within the education sector.

The policy documents since 2015 have encompassed the full incorporation of an ESIS, implementation of open education programs, continued pursuit of a comprehensive lifelong education program based on ICT, the full transition of government services to an electronic format, achievement of full literacy within the population, and the establishment of general education instructor competency standards.

These factors demonstrate that the Mongolian government has established the foundation for the provision of quality education based on ICT through numerous programs and projects.

The "Electronic Mongolia" national program [1] was approved in 2005 and the goal of incorporating ICT into every sector was put forward. As a result, ICT development in Mongolia was ranked at 14th among the 34 Asian countries that were assessed in 2017 as seen in Fig. 1. [2].

ICT BASE INDICATORS (2018)

The foundation for the Mongolian Communication and Information Technology Agency was first established in 1921. ICT is currently seeing rapid development across the world and is playing a significant role in the social, economic, human development, personal freedom environment in many countries.

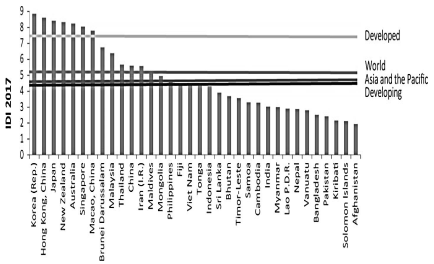
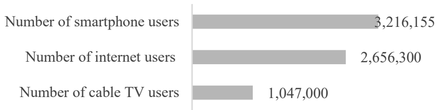


Fig.1: ICT development [2]

Since 2000, countries across have been seeking to establish a society based on ICT in order to accelerate economic and



social development.

Fig. 2: Use of population technology [3]

If we look at the use of technology within Mongolia, we can see that the number of smartphone users in the country in relation to the population is extremely high and stood at 3,216,155 as of 2019. If we look at other indicators, as of 2018, 93% of households had a TV and 31% of households had a laptop.

The Education 2030: Incheon Declaration and Framework of Action [4] establishes that ICT plays a significant role in offering higher quality and equitable education framework, knowledge acquisition, and information retention. ICT is also seen as a key factor in driving forward many objectives such as SDG 4.4 (core and ICT competencies), SDG 4.5 (ICT and technological infarstructure in gender equality), SDG 4.6 (using ICT to expand literacy), SDG 4.a (incorperating ICT into the learning environment) and SDG 4.b (ICT competent instructors).

TABLE I. THE CONTENT OF THE TERM

Objective	Description
Enhancing outcomes	Enhancing ICT infarstructure in order to improve education outcomes through equity.
Enhancing education quality	Establishing a teaching and learning environment through the use of quality online resources and methodology.
Enhancing management	Reinforcing ESIS system in order to enhance the efficiency and efficacy of education management.

Incorporation of ICT into education

ICT incorporation in Early Childhood Education

As of 2017, 78.7% of Early Childhood Education (ECE) institutions were connected to the internet and there was a computer for every 1.2 instructors. In addition to providing internet connection and equipping instructors with computers, significant efforts have also been made to put CCTV cameras in ECE establishments and the ECE Institution Standards [5] provide for equipment and technology standards in ECE.

ICT incorporation in General Education

ICT has been recognized as an essential element in producing individuals with 21st century competencies through student-centered, innovative teaching approaches.

Mongolian legislation on general education encompasses the following provisions:

- Foreign language and ICT classes can be incorporated into the elementary education curriculum;

- Native and foreign language, mathematics, ICT, science, jurisprudence classes shall be provided at the secondary education level within the framework of providing independent living competencies to students;

- The high school curriculum shall encompass environmental studies, humanities, social studies, mathematics, and ICT.

The "Computer for Every Child" program was initiated in 2008 and 11,813 XO computers were distributed by 2012 [6]. Moreover, 25,000 laptops were provided to instructors through the "Correct Mongolian Child" program.

The 2006-2015 Education Master Plan established the incorporation of ICT into education as the foremost priority and encompassed the following objectives:

1) Providing internet connection to all schools;
2) Encompassing instructors in ICT training programs;
3) Equipping classrooms with the capacity to carry out distance education activities;
4) Utilization of monitoring, assessment and evaluation software
5) Establishing a national distance education network of schools and regional centers.

Within the framework of the "New Century Education" project funded by a Republic of China loan, all general education institutions in the nation were connected to the interned by the end of 2017. Furthermore:

- 36,340 laptops were provided to 25,000 general education and ECE instructors with 625 smart boards also being provided to schools and kindergartens.
- 27 network centers with GP ON technology were established on existing public infrastructure.

- 122 schools in Ulaanbaatar, 69 schools in province centers and 350 regional schools were equipped with GPON wire cable technology.
- The Ministry of Education and Science (MOES) funds the internet operating costs.

These initiatives and achievements played an essential role in minimizing the geographical distance and differences between regional and urban schools. Furthermore, the integration of technology into teaching and learning is a fundamental step towards enhancing student ICT competency.

As of 2017, at the general education level, the number of students per computer was 12,5 and the number of instructors per computer was 1.6. There is currently no studies into the internet speed, access, maintenance, and technical specifications of the computers. However there is significant evidence of public-private partnership initiatives to enhance internet speed [6]. In January 2017, the first instance of public-private partnership manifested itself in Huawei and Unitel providing a wireless internet solution for regional households and schools through the “House Internet” initiative. The initiative increased the internet efficiency at 35 regional schools and provided further opportunities for distance education and e-content creation.

The Education 2030 Incheon Declaration and Framework of Action clearly establishes ICT as a key resource in providing equal access to quality education. There is however a need to look at inclusivity from many factors such as location, sex, income level and developmental difficulties. Additionally, there is a need to be aware of the social, economic and information differences digital inequalities may lead to.

Despite its benefits, it is worth noting that ICT infrastructure is not the solution to all the issues facing general education but simply a tool through which to approach the solutions to those issues. There is also a need for comprehensive assessment of the quality and implementation of the initiatives in order to provide appropriate technical support. The experiences of developing and developed countries in the region demonstrates that it is challenging to equip schools with high-quality internet given budget limitations.

Higher Education Institutions and ICT

There were 155,248 students pursuing Higher Education in 95 MOES licensed Universities in 2019. Approximately 40% of the Universities train specialists in ICT and 3,606 (2.3% of students) were engaged in this line of training. An MOES study found that students in ICT fields gradually dropped over the last 5 years. Accordingly, there is a significant need for ICT human resources. At a time where the world is moving toward more reliance on ICT and digital technologies, the need for ICT specialists is predominant in not only Mongolia but across the world. When it comes to Mongolia, it is imperative that we develop competent specialists in order to further the development and achievement within the ICT sector and to create an informed society.

Therefore, there is a need for a comprehensive comparative study into the industry demand for competent specialists and the demand for ICT graduates in the labor market. Furthermore, there is also a dire need for studies into increasing enrollment in ICT programs, enhancing the quality of education, increasing the number of competent ICT specialists and incorporating international standards into ICT training.

As of 2018, there were 37 Mongolian students pursuing higher education in ICT in the USA, 3 students in England, 1 student in Australia, 2 students in Canada, 8 students in China, 7 students in Hungary and 17 students in Russia. It is worth noting that these numbers only encompass students sponsored by MOES and don’t include students pursuing degrees through alternative scholarship and financing options.

The establishment of distance education centers not only allowed for University students and instructors to be encompassed in distance education initiatives but also supported the digitization of learning resources and the implementation of distance assessments.

General education instructor ICT professional development initiatives

In addition to the provision of ICT equipment, MOES also undertook efforts to encompass instructors in ICT professional development activities [6].

The Institute for Teachers’ Professional Development (ITPD) initiated the “http://teacher.itpd.mn” portal that not only provides support in a number of key areas but also allows for experience exchange among instructors. According to MOES, there were 21,000 pieces of content on the portal and 26,857 users as of 2019 [6].

The Mongolian State University of Education (MSUE) implements the annual “Multimedia software incorporation into training” competition among instructors and implements distance training initiatives for the Lifelong Education Centers of the 21 provinces. ITPD also developed the “Advanced instructional technology database” “http://urchadvar.itpd.mn” that currently has 985 class recordings.

Instructors with 1, 5 and 10 years of experience are required to undertake 15 days of online training prior to the academic year and 30 days of online training after the academic year. The content of the trainings is determined by instructor request. In 2017, 11,095 instructors were encompassed in the training and it is estimated that 38% of them managed to increase their ICT competency [6].

Within the framework of establishing instructor ICT competency standards, an instructor readiness survey was conducted among 1,917 instructors and it was determined that the preparation of multimedia learning resources and utilizing ICT to learn new and innovative methods were the two key areas of need. Only 14% of instructors responded that they significantly incorporate ICT into their classes. Furthermore, a disparity in ICT competency among new and tenured instructors was also discovered [6].

In order to address these issues, MOES has paid significant attention to the development and implementation of comprehensive instructor ICT standards with the support of UNESCO. These standards have been developed at the core, advanced and specialized level for elementary and secondary education instructors. The 2019 Education

Country Basic Report also pointed at the urgency of developing instructor ICT competency standards for ECE instructors. This need is based on the pursuit of consistency between the use of technologies for young children and the education delivery methods employed by their instructors.

The establishment of core competency standards will allow for the pursuit of education goals, establishment of a unified understanding between instructors, students and other stakeholders as well as consistence within instructor development programs, learning environments, and instructor professional development initiatives.

Quantifiable and consistent ICT competency standard indicators and additional materials (resources such as TPACK, SAMR) will provide instructors with clear and consistent directions on the effective incorporation of ICT into education.

Countries in the region such as Australia, China, the Phillipines, South Korea and Singapore either incorporate instructor ICT competencies into the instructor core copmetencies or establish seperate instructor ICT competencies. Mongolia is currently in the middle of its first attempt to institutionalize instructor ICT standards. The pursuit of these standards are based on the recognition of its potential impact on the continuous professional development of instructors. Cognizant of the significance of the effort, the experiences of the above mentioned countries have been thoroughly studied and numerous government agencies are currently working towards the development of the standards.

There is also great value in looking into the experiences of countries such as Myanmar, Nepal, the Philippines and Uzbekistan where UNESCO implemented projects to establish similar standards.

One initiative that stands out was the “Utilizing ICT to enhance the quality of elementary education in Mongolian regional schools” project implemented by the Tokyo University of Technology between 2012 - 2017. The Tokyo University of Technology and the MSUE worked with a large number of public institutions in order to provide training on the use of ICT and student-centered education as well as the preparation of digital learning resources for a variety of disciplines for 340 elementary and secondary education instructors in 5 provinces. It is worth further researching this experience and developing methods to extrapolate its outcomes across the nation [7].

It is recommended that the Mongolian government continue its efforts to accelerate teaching innovation through ICT, establish instructor core competency standards and enhance instructor training programs and instructor professional development initiatives.

EDUCATION SECTOR INFORMATION SYSTEM

The Mongolian government have placed a significant emphasis on the incorporation of ICT in education as a means to increase equal access to quality education. Some of the programs towards this pursuit include “The ICT in the Education Sector Policy” (2012 – 2016), “Education Sector Information System” (ESIS, 2013), “Development of open content and programs in education [8].

The following activities were carried out within the framework of establishing an ESIS designed to create a unified information database that allows for analysis and evaluation at the management level so as to increase decision making and planning capacity:

- Establishment of a unified ESIS
- Incorporation of the ESIS into education institutions at all levels
- Provision of accurate, actionable information that allows for effective research activities
- Innovating and enhancing the system, establishing the conditions for its sustainable operation

In developing the ESIS particular attention was paid to ensuring that education stakeholders were provided with the capacity to determine their needs, produce and acquire primary source information at a single destination.

Activities to incorporate general education statistics into the ESIS was initiated in 2014. However, the incorporation of ECE statistics was tested in 2017 and initiated in 2018. Currently 1,354 kindergartens, 256,720 students, 29,424 instructors and 25,639 staff have been registered in the system.

The introduction of ESIS has allowed for education related information at the ECE and general education level as well as related human resources, financial management and statistical information to be consolidated into one platform.

An ESIS for the Higher Education sector was developed in the 2018/19 academic year in order to enhance the operations, information flow, monitoring and quality assurance within the sector and support decision making. The establishment of this single system has allowed for more efficient sector-wide planning and decision making as well as increased information transparency for the public.

The incorporation of international statistical methodologies has allowed for the determination of education indicators in consistence with international standards. Mongolia began reporting education statistics to UNESCO in 2002 and this has allowed for Mongolian education statistics to be compared to indicators across the world.

The incorporation of ICT into education statistics has allowed for an information database based on child registration to be established. Additionally, it has facilitated the creation of a single information base that establishes favorable conditions for evidence-based planning and its effective monitoring and evaluation.

Through the establishment of an ESIS in a phased manner, information flow between different levels of education has been enhanced, education operation interruptions have been minimized and financial mechanisms have been revitalized.

Through the connection of ESIS to other systems, the development of other sectors has also been accelerated.

Despite these advances, there is a need to analyze the SDG indicators, determine current progress towards them and determine existing issues through the ESIS

NEW CHALLENGES FACING THE WORLD

In response to the Coronavirus pandemic encompassing the world, the Mongolian special commission decided during its 24.01.2020 meeting that the activities of educational institutions and training centers at all levels shall be suspended from 27.01.2020 to 27.03.2020. This suspension would be extended a further three times in an attempt to protect Mongolians from a potential outbreak of the Coronavirus.

In these unprecedented times, MOES and MIER developed a short-term plan for delivering ECE, secondary education and higher education content to students without interruption. Furthermore, professional agencies, educational institutions and instructors are working with other stakeholders in the following manner.

MOES (Ministry of Education and Science)

Oversaw the development and delivery of ECE, elementary and secondary education content in an online manner while also providing the opportunity to access tele-classes anytime by placing them on econtent.edu.mn.

- Coordinated activities required to implement the suspension of the General Entrance Exam for High School graduates and the removal of the Mongolian language exam requirement for students entering Universities and colleges.
- Given the absence of distance education funding in the 2020 budget, tele-classes were financed through moderate cutbacks in school maintenance expenses (e.g. electricity, utilities and heating) and the redistribution of expenses

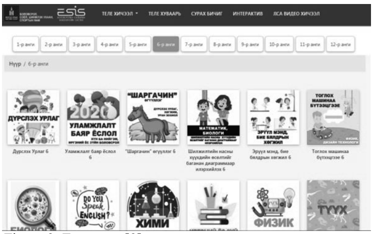


Figure 3. E-content [9]

- Funding previously allocated to student lunches and breaks were redirected to cover the salaries of instructors and specialists working on Saturdays and the procurement of equipment to disinfect their workplace

MIER (Mongolian Institute for Educational Research)

- Developed the content and schedule of tele-classes
- Developed and approved content and schedule of tele-classes for subjects encompassed in the General Entrance Exam for High School graduates
- Consulted on and approved the content and schedule of special needs education tele-classes
- Developed a manual for leading "Instructor Journals" and disseminated it to instructors and educational institutions.
- Organized tele-classes and provided methodological advice to instructors

ITPD (Institute of Teacher Professional Development)

- Organized compulsory professional development trainings for instructors with 1, 5 and 10 years of experience in an online manner
- Provided methodological advice for instructors involved in tele-classes

Education Evaluation Center

- Provided methodological advice to instructors involved in General Entrance Exam preparation tele-classes.
- Copied General Entrance Exam preparation tele-classes to DVD and delivered them to all students

National Lifelong Education Center

- Provided advice to instructors involved in lifelong education tele-classes

Regional Culture and Education Departments

- Worked with educational institutions, instructors and parents to acquire information regarding tele-classes
- Compiled information on student participation in tele-classes and provided reports to MOES

School Administrators, Instructors

- 450 secondary education instructors were involved in the organization of tele-classes. Moreover, 28 special needs instructors and 20 sign language interpreters participated in the development of tele-classes
- School administrators in areas without stable access to television and internet that are mainly attended by children of herders made alternative arrangements at the regional level and provided students with accurate information
- Instructors carried out activities to reinforce tele-classes and evaluate learning outcomes. Instructors, students and parents communicated with each other and other education stakeholders through Facebook. Some educational institutions utilized platforms such as Google Drive, Google Form, Google Meet and Zoom to communicate with students and parents
- Arrangements were made for educators to visit children of herders to become acquainted with class progress and to distribute workbooks, textbooks and notebooks
- In order to provide the opportunity for Kazakh children to obtain education in their mother tongue, the Culture and Education Department of Bayan – Ulgii province translated 581 tele-classes into the Kazakh language and disseminated it to the public

Early Childhood Education

- 20 instructors worked to produce ECE tele-classes

Television Union

- The tele-classes were disseminated through the 16 member networks of the Mongolian Television

Union with approximately 170 television network employees participating in the production of tele-classes

Communications Regulatory Commission, Communication Information Technology Agency

- Phone carries Mobicom and Skytel as well as the information technology company Univision made access to the econtent.edu.mn where tele-classes were placed free of charge and engaged in further activities such as providing phone numbers with free data for students of high school graduating classes and placing tele-classes on their video library
- Distance system connections from mobile phones were facilitated free of charge

NGO, Civic Society and parents

- Participated as guest instructors in tele-classes.
- Supported students by establishing a feedback mechanism with instructors and educational institutions

The following 4 areas have been highlighted as priorities following the outbreak of Covid-19 that has devastated the world.

1. Provision of education content in a tele-class format.
2. Recommendations on activities to reduce the knowledge gap that resulted from the transition to tele-classes and to support the implementation of the curriculum at the beginning of the 2020/21 academic year.
3. Development of an online class platform
4. Monitoring attitudes towards tele-classes and conducting a study into student psyche

1. *Tele-classes*

During the 2020 academic year, Mongolian general education students studied 85-90 of the 160-165 academic days in the classroom and 75 days of education content was delivered through tele-classes. During the lockdown between February 3rd 2020 and March 29th 2020, 4535 tele-classes were delivered through 16 TV channels

The following measures were implemented in order to generate further interest in tele-classes:

- Reinforcement of previously acquired knowledge
- Reducing the knowledge gap resulting from the transition to tele-classes
- Introduction to Mongolian culture and traditions
- Integration of content and methodology, using guest instructors
- Pursuit of equal access to education and making relevant arrangements for children with special needs
- Translation of tele-classes to sign language
- Organization of training for General Entrance Exams for graduating students
- Organization of lifelong education tele-classes for the public

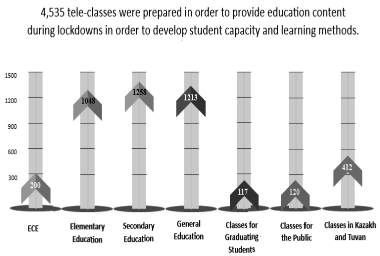


Fig. 4: TV lessons by each level of education [9]

2. *Reccomendations on supporting the curriculum*

A MIER survey of 12,000 parents and 5,700 students found that 46,8% of students consistently watched tele-classes and 30,52% of students sporadically watched tele-classes. [1] This demonstrated a need for measures to reduce the knowledge gap created from the transition to tele-classes. Therefore, beginning in September 2020, kindergarten and general education establishments were directed to spend 1 month reinforcing the content of the previous academic year.

3. *Online class platform development*

Students had the opportunity to undertake their classes through smartphones, tablets, laptops and computers during lockdown.

In order to increase the efficiency of future distance education initiatives, a working group on the establishment of an online class platform that allows for students to ask questions, discuss topics, conduct self-assessment, connect with their instructor and use key learning resources was created.

4. *Monitoring student attitudes towards tele-classes and conducting a study into student psyche*

MIER, MSUE and the IRIM research institute conducted studies into student engagement in tele-classes, tele-class quality, consistence of learning outcomes, instructor and student workload as well as the impact of tele-classes on student psyche. This research would generate the following findings:

- Following the announcement of lockdown, appropriate organization and financial decisions for distance education initiatives were made in a timely manner [10].
- 46.8% of students consistently watched tele-classes and 30,52% sporadically watched tele-classes [11]. 20% of students did not watch any tele-classes. 5% of those students did not have the technical equipment to watch tele-classes and 22% did not have stable internet access [10].
- Tele-class evaluations suggest that the classes were conducted in an above-average to good level. The content was assessed as appropriate and student evaluations returned slightly lower results than

expert and instructor evaluations [10]. An MSUE also demonstrated concerning indicators regarding the impact of lockdown and tele-classes on student psyche [12].

COVID AND HIGHER EDUCATION INSTITUTIONS

Following Covid-19, there will be three distinct lasting implications on Higher Education Institutions (HEI). Firstly, Covid-19 forced the transition to distance education initiatives and the situation served as a golden opportunity for champions of digital technology. In addition to education activities, related activities such as graduation and enrollment shifted to an online manner.

HEIs began to undertake initiatives to enhance their technology base in order to provide equal access to quality distance education.

Secondly, the pandemic can be seen as an example of a monumental but rare tragedy and served as an opportunity to assess the preparedness of HEIs. HEIs began implementing Emergency Management Strategic Plans regardless of their history and reputation.

Finally, the situation provided an opportunity to undertake accurate assessment of Higher Education.

Pressing challenges:

- Students in the regions and rural areas have limited access to video and audio class content due to the high price of data and lack of internet access. This creates a pressing issue in the continued implementation of distance education initiatives.
- The differences in the computers and equipment used by students is creating further discrepancies.
- There are instances of children from herder families living in Ulaanbaatar having a lack of access to internet.

MIER conducted a study into the implementation of distance education initiatives in the higher education sector that returned the following results:

The majority of HEIs provided its instructors with seminars, trainings and comprehensive manuals on the development of distance education content.

One pressing issue in the realm of policy is the general

lack of a distance education legal environment. Existing education legislation also lacks any regulation on the implementation of distance education initiatives. There is a need to establish a unified Higher Education platform and the sharing of quality instructional content. 20% of HEI students did not participate in distance education initiatives. Some of the reasons for this included lack of access to stable internet. Schools, its departments and other stakeholders such as mobile carriers are currently planning initiatives to enhance student engagement in distance education classes.

CONCLUSIONS

The following actions need to be considerable:

1. In determining the effectiveness of ICT in education projects and programs, it is worth assessing tangible outcomes rather than simply looking at surface factors such as the number of computers and internet speed.
2. There is a need to provide assessment equipment and carry out effective monitoring and evaluation activities into the use of ICT.
3. There is a need to research the use of ICT in lifelong education in order to provide equal access to quality education.
4. There is a need to enhance the content of current instructor training programs through ICT, support ICT based innovation, and to support the constant professional development of instructors.
5. Mongolia needs to study experiences such as Australia's implementation directives, Bangladesh's instructor content and experience exchange portal, South Korea's self-evaluation resources, the Phillipines' instructor development programs, Singapore's professional development programs based on the needs at different education levels, and Uzbekistan's outcome-based mechanism but also needs to develop its own unique methodology.
6. There is a need to implement ICT based education in an accurate, flexible manner in order to meet the needs of herder families who pursue a transient life.
7. There is a need for a national policy to support a heutagogy approach to independent education based on ICT.

REFERENCES

[1] Government Resolution No 216, [online]. available: https://legalinfo.mn/mn/detail/12120, 2016.
[2] ITU. Measuring Information Society Report , 2017. pp. 73-109.
[3] Communications Regulatory Commission, 2019 core indicators, [online]. available: www.1212.mn, 2019.
[4] UNESCO, “Mongolia, Education Policy Review”, [online]. available: https://events.unesco.org/, Ulaanbaatar, 2019.
[5] Agency for Standard and Metrology, [online]. available: https://www.masm.gov.mn/, 2015.
[6] MIER, “Country Basic Report”, Ulaanbaatar, 2019, pp. 262-280.
[7] Yamaguchi-Takada-Lab, [online]. available: http://www.yamaguchi.gsic.titech.ac.jp/?page_id=4615, 2017.
[8] Mongolian Parliament Directive 12. [online]. available: https://legalinfo.mn/mn/law?page=law&cate=33&active=1&page=1 &sort=enacteddate&page=1. 2020
[9] MOES & MIER, “ General education curriculum implementation and factors that affecting it”, survey report, Ulaanbaatar, 2019, pp. 28-54
[10] MOES & IRIM, “Tele-classes, outcomes and its consistence”. [online]. available: http://www.usof.gov.mn/, Ulaanbaatar, 2020.
[11] MIER, “Tele-class engagement survey”, [online]. available: http://www.mier.mn/wp-content/, Ulaanbaatar, 2020.
[12] MSUE, “Impact of lockdown on student psyche”, Ulaanbaatar, 2020, p.149.
[13] “Development of a Campus University”, [online]. available: https://must.edu.mn/mn/content/t/1?page=233. Ulaanbaatar, 2010.
[14] “Higher Education Reform Roadmap”, report, [online]. available: https://www.num.edu.mn/files/.pdf, Ulaanbaatar, 2010.
[15] ICT core indicators. [online]. available: http://crc.gov.mn/k/2ZV, 2018.
[16] ITPD, “Education for Sustainable Development project, Instructor professional development needs assessment”, Ulaanbaatar, 2016. pp. 87-104.
[17] ITPD, “Advanced Education Technology”, [online]. available: htpp:// urchadvar.ipd.mn. 2019.

[18] Mongolian Parliament: Labor Legislation, [online]. available: https://legalinfo.mn/mn/detail/565. Ulaanbaatar, 1999.

[19] Mongolian Parliament: Elementary and General Education Legislation, [online]. available: https://legalinfo.mn/mn/detail/9020. Ulaanbaatar, 2002.

[20] Mongolian Parliament: Education Legislation, [online]. available: https://legalinfo.mn/mn/detail/565. Ulaanbaatar, 2002.

[21] Mongolian Parliament: Higher Education Legislation, [online]. available: https://legalinfo.mn/mn/detail/251. Ulaanbaatar, 2002.

[22] Mongolian Parliament: Early Childhood Education Legislation, [online]. available: https://legalinfo.mn/mn/detail/9020. Ulaanbaatar, 2002.

[23] MOES, Academy of Science, “Incorporation of Education for Sustainable Development principles into general education curriculum”., Ulaanbaatar, Soyombo Printing LLC, 2011. pp. 39-97.

[24] MOES, Electronic Content, [online]. available: http://econtent.edu.mn. 2013.

[25] MOES, Academy of Science, “Education for Sustainable Development in Mongolia: Multi-level research”, Ulaanbaatar, 2014. pp. 26-85.

[26] MOES, “Education Sector Information System”, [online]. available: htpp: //esis.edu.mn/, 2017.

[27] MOES, “MOES Operations Report”, Ulaanbaatar, 2017, pp. 84-97.

[28] MOES, “ MOES 2006 – 2019 Statistics”, [online]. available: https://www.meds.gov.mn/, Ulaanbaatar, 2019.

[29] MOES, ITPD, Academy of Science, “Methods to incorporate Education for Sustainable Development into general education and general education support activities”., Ulaanbaatar, 2018, pp.35-51.

[30] MSUE, “ Multimedia software incorporation into training” 11th national competition. Ulaanbaatar, 2019, pp. 58-73.

[31] M.Munkhbaatar, O.Dunjinnamdag, “ICT in Education”, Ulaanbaatar, 2017, pp.28-43.

[32] “Sustainable Development Policy – 2030”, [online]. available: https://sdg.1212.mn/, Ulaanbaatar, 2016.

[33] Yu. Altantuya, “ Incorporation of Education Sector Information System into the Mongolian Education Sector”, Ulaanbaatar, 2018, pp.28-59.

Cluster Analysis on Mathematical Ability of the Children

Narantuya Jamyandorj
Department of Science and Engineering
Mandakh University
Ulaanbaatar, Mongolia
narantuya@mandakh.edu.mn

Ganbat Tsend
School of Information and Communication Technology,
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
ganbat@must.edu.mn

***Abstract*— Mongolian territory is 1,5 million square kilometers, there are 1.439 pre-schools that including 263.333 children. Cluster analyses on mathematical ability of pre-school children is made on selected one province and city according to the Mongolian regional zone. According to the research, totally 450 children between age of 3 - 5 years old (30 children at each age 3, 4 & 5) performed task with 5 questions. The purpose of the research is classification of province zone, determination of inequality and difference between rural and urban area. It is made support for developer's policy and decision makers of education under the base of financing and sharing kindergarten budget, specialization, retraining of teachers and children, developing, elaboration and planning curriculum.**

Keywords*— *data mining; clustering; HCPC; R programming; data visualization;

I. INTRODUCTION

Strategy of Mongolian regional development concept 2010. According to the new general plan of Ulaanbaatar approved by Government, it will be developed as an independent region. It said that the connections of society, culture, economy and institutions for the interprovince will be improved [1].

It was approved assessment procedure of development progress, ability and skills for pre-school children [2]. It is intended to study mathematical ability by age classification under the base of selected cities and provinces from regional zone [3].

Fig. 1. Four years old participant who perform tasks

II. RESEARCH DATA COLLECTION

1.)
"Policy of Mongolian regional development" [1].

Fig. 2. Mongolian regional development concept.

years old children of selected provinces and city who participated in the research

TABLE I. CHOSEN PARTICIPANTS

Age, month	Bayan – ulgii	Dornod	Umnu-Govi	Khuv -sgul	UB-SB District	Total
3 ages	30	30	30	30	30	150
4 ages	30	30	30	30	30	150
5 age	30	30	30	30	30	150
total						450

The datas were collected by mail and online from teachers of pre-schools from selected regions (Fig. 2.) Due to COVID-19, all education system has transferred to online.

The research has started on the 7 January 2019, Mathematics was selected from the pre-school education curriculum following "Music and Fine arts", "Mathematics", "Nature and Social environment", "Language", "Movement and Health" and "Development of Socializing".

Research data were based on performance of the children according to 5 tasks of the Mathematics [8]. Following are: (Table I.)

Task 1: sorting one and plural
Task 2: sorting items by size and naming
Task 3: sorting items by similarity, nominating and connecting
Task 4: using and recognizing position words
Task 5: using and recognizing time expressions: now and after

Task evaluation: it was evaluated such as "very good- 5 scores", "good- 4 scores", "satisfied- 3 score", "bad- 3 score", "very bad- 3 score" (Table II.).

The survey wasn't done due to corona, schools and pre-schools were locked down.

Each child performance of the tasks 1-5.

Tasks 1-5 are indicated such as: q1, q2, q3, q4, q5.

TABLE II. INITIAL 10 COMPONENTS OF BAYAN-ULGII PROVINCE DUE TO SURVEY

kidID	age, month	q1	q2	q3	q4	q5	Aimag
001	3.5	4	3	3	4	2	Bayn-Ulgii
002	3.5	3	2	3	2	2	Bayn-Ulgii
003	3.11	3	2	3	2	2	Bayn-Ulgii
004	3.4	2	1	2	1	1	Bayn-Ulgii
005	3.7	4	2	3	3	4	Bayn-Ulgii
006	3.2	3	3	4	3	2	Bayn-Ulgii
007	3.3	4	3	4	3	3	Bayn-Ulgii
008	3.1	4	3	4	3	3	Bayn-Ulgii
009	3.9	2	1	2	2	1	Bayn-Ulgii
010	3.2	2	1	3	1	1	Bayn-Ulgii

III. METHODOLOGY

Analysis of the Data mining of the Hierarchical Clustering on Principal Components – HCPC is possible to calculate by algorithms of methods PCA, MCA, MFA in program R. [5, 6].

Evaluation mathematical ability of the children is measured by next four formulas.

There are as follows, where $|p - p'|$ is the distance between two objects or points, *p and p'*; *mi* are the mean for cluster, *Ci*; and *ni* is the number of objects in *Ci*.

Minimum distance:

$$dist(c_i, c_j) \min_{p \in c_i, p' \in c_j} \{|p - p'|\}_{min} \quad (1)$$

Maximum distance:

$$dist(c_i, c_j) \max_{p \in c_i, p' \in c_j} \{|p - p'|\}_{max} \quad (2)$$

Mean distance:

$$dist_{mean}(c_i, c_j) = |m_i - m_j| \quad (3)$$

Average distance:

$$dist_{min}(Ci, Cj) = \frac{1}{n_i n_j} \sum_{p \epsilon c_i, p \epsilon c_j} |p - p'| \quad (4)$$

Hierarchical Clustering on Principal Components

The formula was made classification of each cluster depending distance to the nearest cluster center.

$$\max_{i \in q} \min_{q' \neq q} d(i, C_{q'}) \quad (5)$$

$C_{q'}$ to cluster center *q'* distance, it is classified maximum and minimum mean of the nearest and farthest distance and different means.

Using this formula and calculation, it is described difference of it result of cluster analysis on age classification for children according to the regional zone.

IV. PROCESSING

Cluster analysis were made on database at 3 years old participants of pre-school from selected provinces and urban zone according to the Mongolian regional zone (Fig. 3). Three divisions of the Cluster analyses on the

Hierarchical Clustering on Principal Components.
It was used fviz_dend() function of the method HCPC [5].

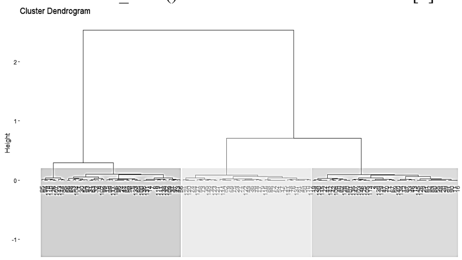
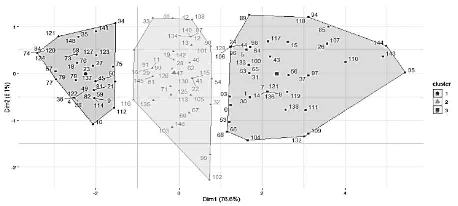


Fig. 3. The division of the cluster analysis involving total 3 years old participants

Hierarchical Clustering on Principal Components are divided into 3 clusters, it is described different means that the nearest and farthest point of the central point. Formula (1-5).

The result of the analysis of the Hierarchical Clustering on Principal Components, using algorithms (MFA) of methods HCPC, it is described different means that the nearest and farthest point of each cluster (Fig. 4).



.

Fig. 4. Total 3 years old participants, each center of the cluster

By (Fig. 4). Total 3 years old children of the cluster task performance are 76.6%, distance to the cluster center is 8.1%.

By (Fig. 5), it is 0.25 mean that green the nearest distance to the cluster center, 0.50 mean is blue middle and 0.75 mean is red far [7]. Algorithm (PCA) of the method HCPC.

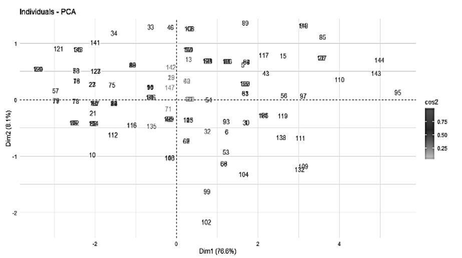


Fig. 5. Total 3 years old participants, tasks 1-5, result of the clusters

The task performance 1-5 for the total 3 years old participants of each cluster is 76.58%, result of each cluster

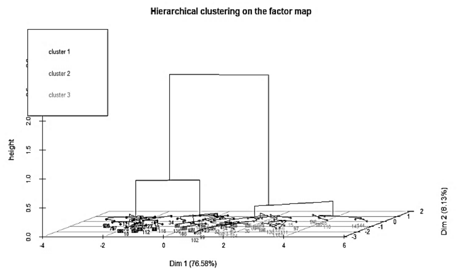


Fig. 6. Total 3 years old participants of each cluster task performance 1-5, result of each cluster

Here (Table III.). Tasks 1-5 of the total 3 years old participants, performance of the clusters (Fig. 6).

TABLE III. TOTAL 3 YEARS OLD PARTICIPANTS, TASK 1-5 RESULT OF THE CLUSTERS

q_id	performance	height
Task 4	0.72%	4.6
Task 1	0.67%	3.1
Task 5	0.65%	2.9
Task 3	0.62%	2.6
Task 2	0.59%	1.9

V. RESULTS

It is described the closest and remote mean of the distance to the center of the 3 years old cluster.

TABLE IV. TOTAL 3 YEARS OLD PARTICIPANTS, TASK RESULT OF THE EACH CLUSTER

Kid-ID	Age, month	Task 1	Task 2	Task 3	Task 4	Task 5	Aimag
Cluster 1 /Close mean to the central distance of the cluster 0.8/							
4	3 years old 4 months	2	1	2	1	1	Bayan – Ulgii
45	3years old 5 months	2	2	2	1	1	Dornod
50	3 years old 2 months	2	2	2	1	1	Dornod
81	3 years old 4 months	2	2	2	1	1	Khuvsgul
137	3 years old 11 months	2	2	2	1	1	Ulaanbaatar, Sukhbaatar
Cluster 1 / remote mean to the central distance of the cluster 3.3/							
74	3 years old 8 months	1	1	1	1	1	Khuvsgul
84	3 years old 1 months	1	1	1	1	1	Khuvsgul
120	3 years old 1 months	1	1	1	1	1	Ulaanbaatar, Sukhbaatar
124	3 years old 6 months	1	1	1	1	1	Ulaanbaatar, Sukhbaatar
121	3 years old 7 months	1	2	1	1	1	Ulaanbaatar, Sukhbaatar
Cluster 2 / Close mean to the central distance of the cluster 0.6/							
3	3 years old 11 months	3	2	3	2	2	Bayan – Ulgii
47	3 years old 8 months	3	2	3	2	2	Dornod
70	3 years old 10 months	3	2	3	2	2	Khuvsgul
129	3 years old 8 months	3	2	3	2	2	Ulaanbaatar, Sukhbaatar
139	3 years old 3 months	3	2	3	2	2	Ulaanbaatar, Sukhbaatar
Cluster 2 / remote mean to the central distance of the cluster 3.1/							
42	3 years old 10 months	3	3	2	2	3	Dornod
46	3 years old 4 months	2	2	2	3	3	Dornod
71	3 years old 9 months	4	2	2	2	2	Khuvsgul
99	3 years old 3 months	4	3	4	2	1	Umnu - Govi
102	3 years old 10 months	3	2	5	3	1	Umnu - Govi
Cluster 3 / Close mean to the central distance of the cluster 0.7/							
7	3 years old 3 months	4	3	4	3	3	Bayan – Ulgii
8	3 years old 10 months	4	3	4	3	3	Bayan – Ulgii
14	3 years old 10 months	4	3	4	3	3	Bayan – Ulgii
131	3 years old 8 months	4	3	4	3	3	Ulaanbaatar, Sukhbaatar
136	3 years old 2 months	4	3	4	3	3	Ulaanbaatar, Sukhbaatar
Cluster 3 / remote mean to the central distance of the cluster 5.6/							
85	3 years old 6 months	4	5	4	3	4	Khuvsgul
95	3 years old 7 months	5	5	5	5	4	Umnu - Govi
110	3 years old 8 months	5	4	4	4	4	Umnu - Govi
143	3 years old 4 months	5	4	5	4	5	Ulaanbaatar, Sukhbaatar
144	3 years old 5 months	5	5	4	5	4	Ulaanbaatar, Sukhbaatar

Here in Table IV seeing 3 years old participants of each cluster task, the cluster task performance 1-5; it has different result task 1 in the cluster 1 mean (1.7, overall mean 2.8).

VI. CONCUSION

Comparing result of Hierarchical Clustering on Principal Components for total 3 years old participants from 3-5 years old participants of the selected district, cities and provinces according to Mongolian regional zone, it was described:

- Ulaanbaatar of central zone that is cluster 2 distance is 0.6 the nearest (Fig. 4) to each cluster center. Cluster 3 is farthest distance 5.6 Umnu-Gobi from central region, Ulaanbaatar from urban zone.

- Total 3 years old participants' who are close to the distance of the clusters performing task 4 is 0.72, farthest distance rate is 4.6.

It is concluded that total 3 years old participants from 3-5 years old participants of the total provinces, cities and districts performed task 4 more efficient. Also, children from Ulaanbaatar and Sukhbaatar district from urban zone, Umnu-Gobi from central region that are selected according to the regional zone, have more ability.

Combination using of training and game program with data visualization are made accuracy of the result.

ACKNOWLEDGMENT

Thahk you for Z.Munkhtuul Senior specialist of the Preschool Education for Education Evaluation Center, for cooperation and advice how to make survey.

REFERENCES

[1] Interactive company. It was developed website, Ulaanbaatar, (2015). law information system: It's consistedconsisted of two parts such as legal acts and research systems. Legal acts are included effective researches, printing, downloading and watching. https://www.legalinfo.mn

[2] Greensoft company. It was developed website, Ulaanbaatar, (2021). Ministry of education and science:It's included innovation of education and science, legal acts, international links and provided open information. https://mecss.gov.mn/news/2582/

[3] J. Batdelger, "A test to determine the general level of mental development of a child", Ulaanbaatar, 1996

[4] J. Han, J. Pei, and M. Kamber, Data mining:concepts and techniques. Elsevier, 2011.

[5] S.-S. t. f. h.-t. d. analysis. HCPC - Hierarchical Clustering on Principal Components: Essentials. Available: http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/117-hcpc-hierarchical-clustering-on-principal-components-essentials/

[6] R Documentation. (2020). HCPC: Hierarchical Clustering on Principle Components (HCPC). Available: https://www.rdocumentation.org/packages/FactoMineR/versions/2.4/topics/HCPC

[7] J. Josse and F. Husson, "Selecting the number of components in principal component analysis using cross-validation approximations," Computational Statistics & Data Analysis, vol. 56, no. 6, pp. 1869-1879, 2012.

[8] Ts. Ganbat, D, Sarangerel, "Human and Computer Interaction", Ulaanbaatar, Mongolia, 2018.

Redefining Blockchain for IoT and Embedded Systems

Nidhee Kamble
Veermata Jijabai Technological Institute
Mumbai, India
ndkamble_b17@ce.vjti.ac.in

***Abstract*—**Instead of directly applying blockchain and distributed ledger systems to IoT topologies, these decentralised data structures linked cryptographically must be reconstructed ground-up from its principles to let IoT fully leverage its features [1] by providing operational details crucial to data collected and transmitted from IoT nodes: contents and validation of a transaction, mechanism of a node joining and publishing on the network, or significance or modification in consensus mechanism. Existing blockchains dominating technological advances at present are configurations that are consequences of specific requirements of the above parameters, and hence should not be taken as the only possible solutions. Blockchain implementations for embedded systems need not be limited to Ethereum and Hyperledger frameworks. Bespoke implementations that are tailored to suit the functional and performance needs of these systems can offer more flexibility and customisation [2]. To facilitate such execution, a conceptual threshold must first be crossed: to see blockchain technology not from a purist perspective but as a development that can be remodelled according to use cases without a computationally intensive consensus mechanism. To preserve integrity, embedded systems working with sensor data can hash the readings along with the metadata, like timestamp, to be sent to a gateway which sends these copies to multiple servers that store or process the data. The gateway can rehash received data to verify its link with the source before transmitting it to servers.

An open blockchain philosophy might not be the best option for IoT considering the performance overheads, which can be minimised greatly using permissioned blockchains instead. A proof of work algorithm hence can be replaced by, for instance, Practical Byzantine Fault Tolerance on Hyperledger Fabric, where nodes can take turns to publish information and append blocks to the chain, with the publisher node cycling sequentially. The published block can be voted as accepted or rejected by the rest of the nodes on the network within a given time frame, failing which the publisher changes. This approach can avoid forks by limiting the publisher to one node, also increasing performance by eliminating the need to solve computational puzzles for validation. If structured appropriately, this can solve bandwidth bottlenecks during signal transmission from nodes that source multiple megabytes of data per second, although the balance between message traffic and delegation of consensus voting power to nodes should be balanced to avoid non-linear scaling. Another alternative could be to use IOTA [3] or Blockchain as a Service (BaaS). However, IOTA might not be fit for battery-powered devices in the network [4].

The features offered by blockchain to protect monetary transactions can also be availed by data recording and transmission - a use case more relevant in embedded systems for IoT. Blockchain can be used for not just distributed, tamper-resistant systems, but also for communication security between IoT nodes for autonomous interaction. Integrity of data transmitted between nodes at each stage can be validated in accordance with protocol established for the blockchain network. This data may be related to financial transactions, but can also be raw or processed data from sensor nodes. Nodes that are endpoints can make decisions about data validity themselves before transmission instead of relying on centrally-controlled mechanisms. These use cases would be suited in systems where different trust relationships exist (trust may be partially present or completely absent [5]) and an objective record of transactions (data or monetary) is required. Examples of such applications might be in different sectors of autonomous vehicle industry [6], embedded systems for residential areas or in agriculture for soil analysis. By having a consistent, tamper-resistant record of this data, it can later be used for deriving patterns. Using blockchain in IoT necessitates high levels of coordination in embedded systems in IoT networks between the chosen implementation and low-level security modules. XAIN [7] intends to provide for this configuration.

Keywords—Blockchain, IoT, Embedded Systems, PBFT, Consensus Mechanism

References

1. Suayb S. Arslan, Raja Jurdak, Jens Jelitto, Bhaskar Krishnamachari, Advancements in distributed ledger technology for Internet of Things, Internet of Things, Volume 9, 2020, 100114, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2019.100114.
2. Blockchain for Embedded Systems - Military embedded: https://militaryembedded.com/comms/communications/blockchain-for-embedded-systems, last accessed 2021/09/19
3. James Brogan, Immanuel Baskaran, Navin Ramachandran, Authenticating Health Activity Data Using Distributed Ledger Technologies, Computational and Structural Biotechnology Journal, Volume 16, 2018, Pages 257-266, ISSN 2001-0370, https://doi.org/10.1016/j.csbj.2018.06.004.
4. Atis Elsts, Efstathios Mitskas, and George Oikonomou. 2018. Distributed Ledger Technology and the Internet of Things: A Feasibility Study. In Proceedings of the 1st Workshop on Blockchain-enabled Networked Sensor Systems (BlockSys'18). Association for Computing Machinery, New York, NY, USA, 7–12. DOI:https://doi.org/10.1145/3282278.3282280
5. Tech Talk: Blockchain and Embedded Systems - Barr Group: https://barrgroup.com/tech-talks/blockchain, last accessed 2021/09/19
6. Kamble N., Gala R., Vijayaraghavan R., Shukla E., Patel D. (2021) Using Blockchain in Autonomous Vehicles. In: Maleh Y., Baddi Y., Alazab M., Tawalbeh L., Romdhani I. (eds) Artificial Intelligence and Blockchain for Future Cybersecurity Applications. Studies in Big Data, vol 90. Springer, Cham. https://doi.org/10.1007/978-3-030-74575-2_15
7. Blockchain for Embedded Systems: combining microcontrollers and blockchain paradigms for increased security - Embedded Computing: https://www.embeddedcomputing.com/technology/ai-machine-learning/blockchain-for-embedded-systems-combining-microcontrollers-and-blockchain-paradigms-for-increased-security, last accessed 2021/09/19

The Study of the Information Culture Competencies of Secondary School Teachers

Nyamsuren Amsarvaa
Science section,
Orkhon Khasu secondary school
Ulaanbaatar, Mongolia
nyamsuren_a@orkhonschool.edu.mn

Sodov Tsedendamba
Teacher development center,
Orkhon university
Ulaanbaatar, Mongolia
sodov_ts@yahoo.com

Abstract— **Information Technology Complex (ITC) of Teachers is one of the important indicators to create educational united condition and space in the country and it is defined by performance success of the information purpose life cycle.**

In this regard, we identified some of the issues that need to be addressed in order to diagnose and improve the current state of information culture skills of secondary school teachers and determined the needs to motivate, gain and use the information culture and implement them in the communication with subjects including students, parents and other teachers and supervisors and the development of training with implementation of professional assignments in accordance with the new requirements.

Keywords— Information culture, teachers, competence of ICT, communication skill, current situation, usage of informational culture, culture level.

I. INTRODUCTION

Information Technology Complex (ITC) of teachers is one of the important indicators to create educational united condition and space in the country and it is defined by performance success of the information purpose life cycle. Some author [1] considers that it is important following principles to set up teacher's ITC:

- professional and pedagogical orientation of ITC
- motivation values in the information
- personal experience formalization in the information
- To provide a zone of nearby development of personality traits and information experiences
- Information experience development and to cultivate it
- To integrate information experience.

Apart from these, it is interesting to note that the expressing, interpreting and anticipating roles of pedagogical phenomenon and processes are included. For example, it says that the formation of teacher's ITC is responsible for forecasting, synthesizing, diagnosing, developing, analyzing, interactive and organizational-technological supporting and training.

In the pedagogical and psychological-pedagogical structure, which provides motivation, cognition, and consideration of the main criteria of personal and operational development of ITC, we provide information, research, materials, techniques, personal characteristics, socio-pedagogical methods and techniques. It is important to consider it comprehensively from a methodology point of view. The reason for this is that the need for modern teachers to use information and communication technology in their professional activities, general pedagogical skills, and research skills arises from the development of today's society. In general, the ICT competence of teachers at all levels is considered to be an important condition for success in the educational space, in addition to the specific conditions of the school and the level of students' learning skills, as the educational process needs to be open and understandable to all participants [2].

In this article, we considered:

- Information culture role in the teachers' individual professional activities
- Diagnosing and studying current situation of information culture of secondary school teachers
- The result (teacher information culture needs, levels, motivation, use of ICT, its applicability, ways to improve information culture, in teacher training and professional development programs of the research is integrated and some issues to improve information culture is considered to be submitted

II. THE MAIN CONCEPT

The need for teachers to diagnose and improve their level of information culture, which is part of the general culture of humankind, and their need for self-knowledge, motivation, mastery and use of information culture has been studied through literature and questionnaires. Our knowledge, skills and joint efforts of teachers and parents are very important in directing the information environment, information flow and creating a legal environment for its activities. Professional pedagogy is an important factor in determining the effectiveness of educational technology, so a teacher is a master of the development of pedagogical activities and the professional acquisition of pedagogical values and resources. The history of vocational schools and pedagogical ideas shows that awareness, constant evaluation, value creation, and the transfer of known ideas and pedagogical technologies to new contexts. Re-examining old skills and appreciating their strengths is a very important and integral part of a professional teacher's pedagogical culture.

The level of teacher information culture (TIC) is knowledge of information, information processes, models and technologies; The ability to develop and analyze information tools in a variety of activities and to use open information systems (the Internet, its services, and resources) in teaching is determined by practice. It is becoming necessary to create an information-intellectual environment by taking the opportunity to create a teacher-site by posting the information the students need on the Web-site; There is a need to constantly assess the comprehensive competencies of students in the field of information culture using computers, local and global networks.

In other words, Information Culture Formation Level is

- Skills to use data and information technology
- Recognized the importance of get information via printed and online version, ability to choose necessary information from the information source, understand the information resource influence, communication skills to ICT
- Information culture is standard of ICT, and the system of knowledge, skill and practice which provides cultural and mental values.

The content of a teacher's information culture (TIC) is related to the holistic ability to master it. It is innovative and interesting to differentiate into three parts: ICT-based professional development, pedagogical activities, student education in the context of informatization of education, essential professional relations, and cooperation between stakeholders in education [4], see Table 1.

TABLE I. CULTURAL CONTENT OF TEACHER INFORMATION

Component	**Content of Teacher's Information Holistic Ability**		
	Professional development in the range of ICT	***Pedagogical activity in the informatization of education***	***Inevitable communication and cooperation of profession***
Worldview	The necessity to develop profession in the range of ICT by integrating official, non-official, and gained by environment education	The necessity to organize training based on modern ICT and inevitable communica-tion	The necessity to conduct the pedagogy cooperation and pedagogy cooperation through network productively based on inevitable communication
Knowledge	Knowledge to provide the activities to improve the ICT profession individually	To conduct activities to study together on network activities of pedagogy	possibility and knowledge about ICT tools beneficially using network
Teacher's experiment of training and development activity	Ability to professionally develop using the methods of official, non-official, and gained by environment education	Ability to work individually and in team with students and other subjects based on ICT tools and social internet	Skill to communicate and cooperate with remote graduates
Communica-tion	To improve the general level of communication culture	Ability to conduct online discussion at the information and education circumstance oriented to implement train and develop purposes	Ability to use network communication in training activities by pedagogy cooperation in inevitable communication principles
Health (ecological)	Ability to keep own body, mind and mentality in the Global Communication Information Society	To conduct activities to protect body, mind and mentality health in the Global Communicatio n Information Society condition.	To create a favorable psychological environment in the pedagogy cooperation network based on the result of inevitable communication

At the same time, it is studied as following 4 parts, it can stimulate teachers' comprehensive IT capabilities motivation (willingness to organize information retrieval and analysis systems, interest in mutually beneficial exchange of information with relevant subjects, habitual use of ICT in pedagogical activities), cognition (basic information retrieval and processing knowledge of algorithms, in-depth knowledge of current telecommunication systems, knowledge of multimedia and interactive technology tools), operational (knowledge of ICT retrieval systems, formal methods of analysis, knowledge of information and communication technologies), personal (analytical thinking, communication and variable quality) methodology and variability).

On this basis, we conducted a survey about diagnosis ITC from 50 secondary teachers about the need for self-awareness and motivation; teacher information culture; Use of ICT; ICT competencies of teachers; diagnosis of teachers' ICT skills; and the ability to communicate with subjects using ICT. The results of the survey are presented in Table 2.

TABLE II. THE RESULTS OF DIAGNOSING EXPERIENCE INFORMATION TECHNOLOGY COMPLEX SKILLS

Cultural component	***Valuable***	***Personal***	***Cognitive***	***Operational and technological***	***Formation and level of informational culture***	***Percent***
Adaptive	14	28	30	12	28	44,8
Reproductive	34	15	5	18	20	36,8
Creative	2	7	15	20	2	18,4
Total	50	50	50	50	50	100

The following conclusions can be drawn from this. At the creative level, 18.4% of the teachers surveyed, at the intermediate level, 36.8%, and at the lower or adaptive level, 44.8%.

According to most researchers:

- education system that determines the general level of intellectual development of people;
- information infrastructure that determines a person's ability to receive, transmit, store and use information;
- The development of the country's economy (computers, television, telecommunications, etc.), which determines the material capabilities of people in the use of modern ICT, are the main factors influencing the development of information culture.

Teachers' information culture can be considered in the context of mental, motivational, courageous, emotional, subject-practical, and self-regulation [5]:

- In the mental frame: knowledge of thinking and information technology
- In the motivation frame: motivation for the development of information culture
- In the courage frame: purposeful activities in the information environment, resistance to the conditions of search and processing of information for educational purposes, steady mastery of new information technologies
- In the emotional frame: the ability to understand personal emotions in the context of information retrieval and processing, the ability to overcome technical and other obstacles in working in an information environment, the ability to realistically assess the level of personal culture and information culture;
- In the subject-practical frame: ability to acquire new knowledge, types and forms of activities in the information environment, readiness to cooperate using information technology, acquisition of operational skills, information processing skills, ability to use information technology, orientation in the information environment
- In the self-regulation frame: the ability to master the technology of searching, transforming and using information, the ability to integrate their activities and the level of information culture with social and professional experience.

Teachers of all levels of schools have an information culture, which expands their teaching methods and changes positively as a result of professional necessity, personality traits, aspirations, positions and research results, and their level of information culture can be determined by the following professional indicators [3,6]:

- Interest in improving the educational process through the use of information technology and the need to constantly update their knowledge in the field
- Cautious and humane approach to news and information objects in the Internet environment, critical and analytical approach to the use of information
- Ethical, realistic and patient approach to computer-network communication
- Democratic and ethical forms in pedagogical communication with others in a virtual environment
- Adaptation to the information society and have professional energy.

Teacher specializes in science, anthropology, pedagogy, and the art of teaching; and First, professor who knows the basics of university teaching theory and the many subjects that students' study. Second, he or she has a broad knowledge of new developments in the professional field and literature. Third, kind, helpful and able to make people interested in science; Fourth, it is very important in today's society to be able to use creative ideas and research methods effectively. In this regard, in any society in addition to formal education, the issue of self-education (SES) has always been important, as it addresses the needs of society, self-awareness, and the cultural, educational, professional, and scientific advancement of intellectual needs. Today's young people like it because it is characterized by free choice of education.

The methodology for using ICT in training is: First, ICT in certain circumstances, and the rules for selecting these components; Second, the planning, editing, selection, and adaptation of ICTs to specific subjects in a specific psychological and pedagogical context; Third is the science of applying technology to achieve the learning objectives in a specific context of psychological-pedagogical content of a subject. Therefore, the training to use this ICT is:

a. Explaining the learning process using ICT

b. Analyzing existing ICTs

c. Introducing ICT in certain pedagogical and psychological contexts

d. Creating and modeling ICTs on their own, it is possible to enrich and develop the methods of their use (to provide a pedagogical basis for studying the realities of pedagogy). Taking into account the specifics of this methodology, the teacher needs to consider the relationship between the content, purpose, principles, methods, forms, activities of educational subjects, and diagnosing and monitoring the learning process when developing the methodology for using ICT in the classroom.

In today's world, where the development of information culture is an important area for the professional training of university teachers, a teacher working with students can only play the role of disseminator of information culture if he/she develops himself/herself [4]. The method of intellectual mapping has been widely used to develop teacher's creative thinking [8]. This approach is ideal for developing a teacher's comprehensive cultural competence, as it is non-linear in its thinking, activates the process of discovering connections, allows for a multifaceted view of the problem, describes the mind and participates in problem solving, and helps self-identification and decision-making. By cultivating all the qualities of a teacher, we become a teacher (a source of professional knowledge, skills, practices, and experience), a counselor (a student who completes assignments online, a counselor for online information resources, and answers questions), and a facilitator in the learning process (to conduct activities, to manage their joint activities, to organize administrative work), motivator (to encourage students to engage in creative activities and to acquire knowledge independently), planner (to plan the form of organizing the learning process and to carry out its activities in the most effective way), coordinator (facilitator of training).

We agree that the 5 sections, in order to develop the ITC for them through the teacher professional development system: first, a creative diagnosis, including diagnostic tools to determine the criteria, indicators, level and level of the formation; Second, action-technology with the principles of educational activities, including the direction of professional education of teachers, the importance of information values, stages of development and methods; Third, the system-causal part, which identifies the organizational and psychological factors of training and creates the necessary conditions; Fourth, the content-activity direction to develop the ITC independently of the teacher through the implementation of the educational content; Fifth, concept-objectives, which include theoretical methods, goals, main ideas, and directions based on the ITC in the professional development system [1].

Socio-demographic mobility caused by mass migration around the world; the process of self-determination of national culture has intensified; Our research has shown that the issue of teacher culture and education cannot be separated from the issue of multicultural education proposed by Prof. S. Tumur-Ochir [7] in connection with the existence of national aggression in society.

III. Results

According to a study diagnosing ITCS in secondary school teachers (Figure 1), most of them (44.8%) are at the ITCS adaptive level, a significant (36.8%) are at the ITCS reproductive level, and a very small percentage (18.4%) are at the creative level. This is due to the low level of information culture of teachers.

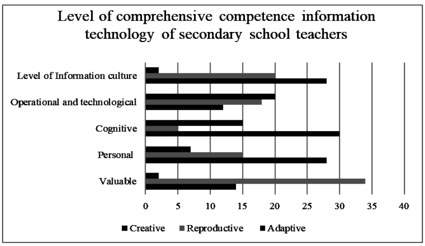


Fig. 1. Level of comprehensive competence information technology of secondary school teachers

This suggests that their motivation to acquire ICTs is dominated by external material motivations and that they are less prepared for self-development. It was also found that there is a discrepancy between the level of knowledge, actions and technology of secondary school teachers, as well as the information and communication technology knowledge of teachers in the secondary school system and the requirements for teachers today.

We, the teachers, believe that it is time to investigate and correct the fact that the ICT knowledge and skills of some students exceed the ICT capabilities of the teacher, which negatively affects the smooth running of the learning process.

A lot of work is being done in our country to modernize learning technology using ICT. One of them is the e-learning seminar on “Transformation Teaching to Learning for High Education” (10) held on September 28-28, 2021. In this workshop, teachers will learn about new technologies, new ideas to use in the classroom, and the needs and requirements for the transition from teaching to learning, learning design and planning, the concept of open education, and the effective use of teaching and learning methods. , there is a need to focus on developing a methodology that can be used creatively in the teaching process, which should be taken as a social order in the Mongolian education sector.

IV. CONCLUSION

1. The first is the ability of the individual to use information and information technology; Second, the ability to communicate with ICTs by selecting what you need from the information resources, realizing the importance of using information in print and electronic form and understanding the conditions that affect the source of the information; Third, an information culture can only be achieved through a comprehensive set of knowledge, skills, and practices that support ICT and human action on moral norms, cultural, and spiritual values.
2. The challenge for the education sector is to develop ICT methodologies and technologies in the curricula of teacher training colleges and in-service teacher training programs, to organize a wide-ranging discussion of teachers' results, and to implement them as a matter of urgency.
3. With the proliferation of “Open Schools” where AI can develop programs, time, and teachers to improve their culture, multiculturalism and creative thinking can be used to enhance comprehensive competencies and skills in the workplace by combining the advantages of traditional technology with the benefits of e-learning. The opportunities for development have increased. In such a situation, there is an opportunity to increase the comprehensive information culture capacity of secondary school teachers in accordance with the e-learning recommendations “Accelerating the transition from teaching to learning”.
4. We, the teachers, need time to systematically implement specific projects and programs that support the creative ideas and motivation of the school administration by separating the teachers from the materialistic mentality of the outside world and creating a creative mindset and a high level of ICT culture. However, it is important to note that if you do not work on it at all, you may end up behind.
5. Create and use an open educational media system that reflects the concept and content of the revised education law related to the psycho-physiological characteristics, living environment, of the student-subject studying online and employer's order, and national development perspective. We are proposing to improve our knowledge by sharing information and experiences.

References

[1] V.L.Akapiev “Model of the formation of information technology competence of teachers”, Electronic scientific publication Almanac Space and Time. T-1. Issue-2, 2012.

[2] T.V. Barakina "ICT Competence of a Teacher as a Condition for Success in Educational Space" collection of articles of international scientific and practical conferences "Information culture of modern childhood" p103-106, Chelyabinsk, 2019.

[3] A. Nyamsuren, Ts. Sodov “Developing an information culture teacher” “International Symposium on Computer Science, Computer engineering and educational technology ISCSET-2020”, TUD press, IBN Scientific Workshop Proceeding, p60-64, Printing in Germany, 2020.

[4] V.A. Polyakova, O.A. Kozlov “Model of training pedagogical and managerial personnel in the field of using information and communication technologies for the system of higher professional education, Pedagogical informatics, No. 2, p.44-57, 2015.

[5] P. Lhagvasuren, Kh. Munkhtsooj, A. Nyamsuren, L. Odonchimeg, B. Purevjal, Ts. Sodov, Kh. Tamir, L. Udval, and Ts. Tsetsegsuren “Information technology means and methods of using them in training (for graduating students and teachers)”, 2021.

[6] A. Nyamsuren, Ts. Sodov “Some issues of personality information culture in the information society”, Ministry of Education, Culture and Science of Mongolia, Mongolian Knowledge Society, Margad Institute, Research Papers Series-VIII, p.107-110, 2021.

[7] S. Tumur-Ochir “Comparative study of world education: on the example of the United States, Great Britain, Germany, France, Japan, Russia, Finland and Mongolia” pp.253-285, 2020

[8] D. Badarch, B. Munkhkhuj “Systems thinking: recognizing diseases in a changing world”, 2014

[9] S. Tumur-Ochir (2020) “General and educational sociology” p.332-351

[10] National e-learning seminar on “Transformation Teaching to Learning for High Education” organized by the Open Education Center of the MUST, the Ministry of Education and Science of Mongolia, Institute of Teacher’s Professional Development, 2021.9.28-30

Towards End-to-End Estimation of Camera Trajectory With Deep Monocular Visual Odometry

Sanketh Rajendra
Department of Electrical Engineering and Information Technology
Chemnitz University of Technology
Chemnitz, Germany
sanketh.rajendra@s2018.tu-chemnitz.de

Shadi Saleh
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
shadi.saleh@informatik.tu-chemnitz.de

Batbayar Battseren
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
batbayar.battseren@informatik.tu-chemnitz.de

Julkar Nine
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
julkar.nine@informatik.tu-chemnitz.de

Wolfram Hardt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
wolfram.hardt@informatik.tu-chemnitz.de

***Abstract*—In the growing world of artificial intelligence (AI), there are tremendous opportunities in computer vision, as today's robots are already equipped with cameras or one can easily be attached. There are a variety of machines and robots moving autonomously in industries or warehouses, or something as new as self-driving cars on the road. Besides that, the localization of the robots is essential, it is also very important to track the path or perform visual odometry (VO). Since most of the cameras in the world are single cameras, this project is intended to work with image sequences from a single camera. This is not the only challenge, but the demands are higher than for stereo vision applications.**

This paper aims towards an end-to-end Monocular Visual Odometry (MVO) with deep learning model Deep-MVO to successfully track the trajectory of a moving camera from input RGB image sequences of a single or monocular camera. A deep Recurrent Neural Network (RCNN) framework model is designed to accept image sequences from video frames and process them to predicted the trajectory of a moving camera and plot it. Since it is trained and deployed in an end-to-end manner, it infers poses directly from a sequence of raw RGB images without requiring inputs from supporting sensors. Extensive experiments with the KITTI VO dataset show good performance compared to state-of-the-art methods.

Keywords— visual odometry, monocular visual odometry, deep-learning, Deep-MVO

I. Introduction

There is a great need for automation in industry, and computer vision plays a major role in this process. In advanced development of autonomous vehicles and autonomous robots, it is very important to track the location and trajectory of the vehicle. Most of these vehicles are equipped with cameras as peripherals. They may perform other various operations such as detecting depth or obstacle avoidance. The data from these cameras can be used to track the trajectory of the camera.

Performing the VO through Deep Learning (DL) is not only for tracking the camera path, but can also be used as a support system for localization. This paper presents a DL method targeted towards end-to-end MVO. As most of the camera setups are monocular, the model is targeted to work on image sequences from a single camera instead of stereo cameras which have an undisputed advantage over monocular cameras. For instances There are numbers of studies [1, 2] been done to perform stereo vision task with monocular cameras.

The most well-known example of VO being used is on the mars rover [3]. This is the best example to explain the importance of MVO. Localization is hard on mars as there is no GPS or milestones, but the mars rover is equipped with good cameras and packs state of the art computing capabilities. NASA performed visual odometry by geometric methods to track the rover's path on mars.

With a DL approach, this project defines a model development towards end-to-end deep learning monocular visual odometry.

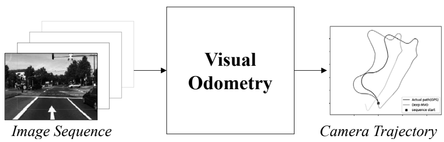


Fig. 1. Monocular Visual Odometry

A. Visual Odometry

Odometry is an estimation of position changes by reading data from various sensors. VO or Egomotion determines the orientation and trajectory of a robot by analyzing the images from the camera physically placed on the robot. VO is used in a variety of applications, such as autonomous robots, unmanned aerial vehicles, etc.

B. Deep-MVO

Deep-MVO is using a DL model to perform VO from a single camera. The architecture of the model used in this project consists of a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN).

C. *Supervised learning*

The supervised learning method is selected for this project as it requires less time to train, and it is generally preferred if ground truth values are known. The ground truth is an actual value of the output that the user provides to the DL model. The ground truth can be any form of result such as actual depth or correct answers for image classification.

II. RESEARCH CHALLENGES

This paper aims to solve MVO by achieving reasonable accuracy in tracking the trajectory of a monocular camera using DL. DL models are improved through the training and testing processes and therefore it evolved to achieve better results. The main challenge is to achieve not only accurate results, but also consistent results so that such models can be used in unknown environments. If a model is trained to achieve high accuracy in a known environment, it may fail in an unknown environment. This paper aims to address this problem as well.

III. LITERATURE REVIEW

To select the best learning method, first the current traditional methods of visual odometry and similar methods have to be analyzed. Visual odometry has been a developing concept for years and different methods and approaches have been established to achieve this concept. These methods have improved and modified a lot over time, and in the case of geometric methods, they are sometimes used as standard libraries.

A. *Geometric methods*

Geometric based methods employ algorithms to compute the path of the camera with equations and mostly a mathematical approach is employed for the whole computation. For instances in [4], a geometric calculation based two alternative approaches were proposed, which are the monocular version VISO2-M and stereo version VISO2-S.

B. *Simultaneous Localization And Mapping (SLAM)*

Simultaneous localization and mapping (SLAM) is a computational problem in computational geometry and robotics that involves creating or updating a map of an unknown area while also keeping track of an agent's position within it [5, 6, 7]. SLAM algorithms are customized to the available resources, so they aren't aiming for perfection, but for operational enforcement. Self-driving cars, unmanned aerial vehicles, autonomous underwater vehicles, planetary rovers and newer domestic robots body use published methods.

The key distinction between VO and SLAM is that VO focuses primarily on local consistency and attempts to incrementally estimate the direction of the camera/ robot pose after pose, as well as likely performing local optimization. SLAM, on the other hand, aims to provide a globally consistent approximation of the camera/robot trajectory and map.

C. *Deep-VO*

Learning-based methods are data-driven approaches that use machine Learning techniques to learn motion models and derive VO from sensor readings without directly applying geometric theory. Deep-VO employs a DL-based approach where parameters are not as clearly defined as in geometric based methods.

Relatively deep-MVO is also likely to be less accurate than geometric methods. But deep-MVO makes up for accuracy in other parameters such as consistency. Often deep-VO is a complement to SLAM. In [8], a DL-based approach is proposed for visually impaired and blind people to navigate and avoid obstacles outdoors using a mobile device.

Table I illustrates the literature survey result. The main comparison factors are the translational and rotational errors. DeepVVO and DeepAVO are DL based methods to track the camera trajectory with monocular camera [9, 10]. The FTMVO uses an iterative 5-point method to estimate the camera motion, and proposed to track low quality features on the ground plane [11]. The SOFT2 approach is the best performing solution on the KITTI Vision Benchmark Suite with 0.53% translational and 0.0009 deg/m rotational error [12].

TABLE I. LITERATURE SURVEY RESULT ON VISUAL ODOMETRY

Method	Type	Translational Error [%]	Rotational Error [deg/m]	Comments
VISO2-S	Geometric	2.44	0.0114	*Library VO*
VISO2-M	Geometric	11.94	0.0234	*Library VO*
DeepAVO	Deep-VO	4.1	0.0125	*DL approach*
DeepVVO	Deep-VO	2.19	0.0088	*DL approach*
FTMVO	Geometric	2.24	0.0049	*Low Rotational error. Slow run time approach*
SOFT2	Geometric	0.53	0.0009	*Best published method*

IV. METHODOLOGY

The aim of this project is to achieve MVO with DL-based approach. The model will be trained and evaluated with sequence of image dataset with actual position ground truth.

The input of the implementation would be an only RGB image sequences from a single camera. There is no other input from other sensors such as camera orientation or camera traveling speed as the aim is to develop an end-to-end learning model. The output of the implementation is a camera travel trajectory in 2D space which will be visualized in the 2D graph.

The proposed concept is composed mainly of a Recurrent Convolutional Neural Network (RCNN). It consists primarily of CNN-based feature extraction and RNN based sequential modelling [13].

A. *RCNN Framework*

The architecture of the proposed end-to-end VO system is shown in Fig. 2, which is based on [13]. It takes a video clip or a monocular image sequence as input. Two consecutive images are stacked together to form a tensor for the RCNN. This image tensor is fed into the CNN to estimate the poses. This is passed on to the RNN for sequential learning. Each of the two consecutive images will result in one pose estimation. This framework develops eventually as the network is trained over time with new images and poses.

The advantage of the RCNN based architecture is to allow simultaneous feature extraction and sequential modelling of VO through the combination of CNN and RNN.

B. Dataset

The KITTI VO/SLAM benchmark consists of 22 image sequences, 11 (Sequence 00 - 10) of which are associated with ground truth. The remaining ten sequences (Sequence 11 - 21) are only given raw sensor data. The first 11 sequences are used to train, validate and test the model. This dataset presents a significant challenge for monocular VO algorithms since it was captured at a low frame rate of 10 fps, while driving in urban areas with many dynamic objects at top speed up to 90 km/h.

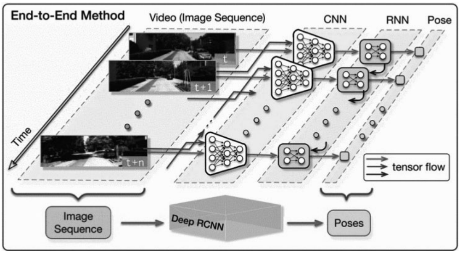


Fig. 2. RCNN Framework [13]

The dataset used to train and test this model is compatible with the model to achieve better results. The images are three channel RGB images. The file type is PNG and the average image resolution of the original dataset is 1220x370 as it varies for each sequence. This was rescaled to a minimum required resolution 184x608 for faster training. There are total of 23634 images used for training and testing.

C. Implementation

The model has been implemented in following order. Firstly, the environment has to be set up for the model to be trained, tested and evaluated. The required main packages are installed in a virtual environment, which are Pytorch 1.8.1 (Torch), torchvision, numpy, pandas, matplotlib, glob2, etc. Once the environment is setup on the Ubuntu machine with a Ryzen 5 processor and 16 GB RAM, it is trained rigorously for 200 epochs in total. For checking the estimation, the 2D trajectories were plotted. Finally, the trained model is evaluated from images sequences. KITTI provides 11 sequences for training, 6 sequences are used for training and the remaining 5 sequences are used for validation. This facilitates the model to work in unknown environments as the training videos are not used for validation. The training process can be saved at any timed and resumed by loading the saved model although training begins from epoch 0.

V. Experimental Analysis and Results

The final trained model was tested with all the 11 sequences of the KITTI dataset. The obtained results are shown in Fig. 3 and Fig. 4. The predicted path is plotted in red (deep-MVO) and the actual path or the data from the GPS (ground truth) is plotted in green.

In Sequence 1 (Fig. 3), a car traveled a simple trajectory. It goes on a fairly unidirectional motion with a hard turn at the end. Even though, they are not aligned due to translational and rotational errors, the predicted path is very close to the ground truth path. In the prediction, the hard turn at the end is estimated correctly, which is important as the model should detect such hard turns in unknown environments. This will be improved with further training of the model.

In Sequence 7 (Fig. 4), the car traveled through a relatively complex trajectory with different turns. In this case, even the predicted trajectory dilated, the most of the turns are estimated correct.

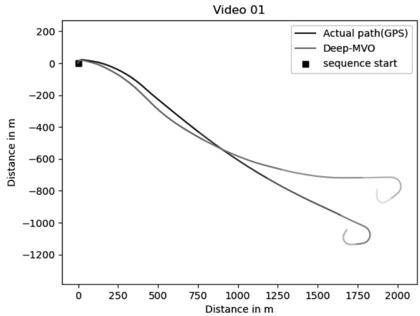


Fig. 3. Prediction result on KITTI Sequence 1

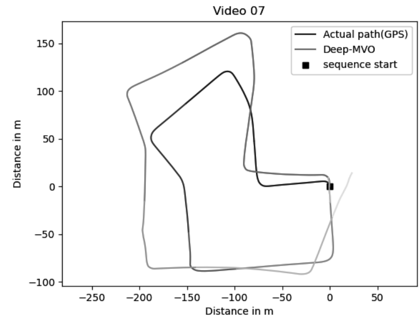


Fig. 4. Prediction result on KITTI Sequence 7

A. Evaluation of obtained results

In other to evaluate and compare the obtained results, there are two main factors were used according to the KITTI evaluation metrics. The mean value of RMSE results from different image sequences is shown in Table II.

1. Average translational RMSE drift (t_{rel} [%]) – This error defines the average point-to-point root mean square error (RMSE) when measuring the total travelled distance by camera
2. Average rotational RMSE drift (r_{rel} [deg/100m]) - This value presents the average point-to-point RMSE of the heading, or to be put in layman terms change in direction from the actual path in deg/100m.

TABLE II. OBTAINED RMSE RESULT

RMSE	Unit	Mean	Highest	Lowest
t_{rel}	%	29.84	44.07	9.8

r_{rel}	deg/100m	0.082	0.172	0.015

B. *Comparisons with monocular visual odometry*

Even though the DL model is no match for geometric stereo vision methods, there is a fair fight in geometric monocular vision methods. Such an example is as shown below in Table III.

TABLE III. COMPARISON WITH MONOCULAR VISUAL ODOMETRY

Sequence	Deep-MVO		VISO2-M	
	t_{rel} *[%]*	r_{rel} *[deg/100m]*	t_{rel} *[%]*	r_{rel} *[deg/100m]*
4	9.8	0.015	0.496	0.0449
5	42.57	0.092	19.22	0.1758
6	44.07	**0.172**	7.30	**0.0614**
7	**11.42**	0.104	**23.61**	0.2911
10	42.50	0.031	41.56	0.3299
Mean	29.84	0.082	19.27	0.1806

According to the translational error results (Table III), VISO2-M approach performed better, except for the Sequence 07, where Deep-MVO achieved better result. On the other hand, the overall rotational error of Deep-MVO was lower than the VISO2-M, expect for the Sequence 06, where VISO2-M performed better. Hence, in order to develop reliable model, consistency should be considered.

C. *Comparison with other DL methods*

DL models are not yet competing with the popular libraries used in geometric methods. Therefore, they are compared with the results of other deep learning models. Table IV shows the comparison with Deep-MVO and Mono-DSO [14].

TABLE IV. COMPARISON WITH OTHER DL MODEL

Sequence	Deep-MVO		Mono-DSO	
	t_{rel} *[%]*	r_{rel} *[deg/100m]*	t_{rel} *[%]*	r_{rel} *[deg/100m]*
4	9.8	0.015	0.82	0.16
5	42.57	0.092	72.6	0.23
6	44.07	**0.172**	42.2	0.20
7	**11.42**	0.104	48.4	0.32
10	42.50	0.031	24.0	0.22
Mean	29.84	0.082	37.6	0.226

The results of the translational and rotation errors in Table IV show that both models perform similarly and there is no clear winner, due to the inconsistency in the comparison. For instances, Deep-MVO obtained better results on the all sequences according to the rotational error. However, according to the translational error, no one outperformed. Nevertheless, the Deep-MVO model performed slightly better than the Mono-DSO model on these datasets, as indicated by the mean error values.

Hence, a consistent model is required to reliably achieve monocular visual odometry through DL methods. These results will improve over time with new environments.

D. *Training and testing*

The model has to be trained with as many as different trajectories to achieve higher accuracy and consistent results to make the model reliable even to work with new video frames in unknown environments in the real world. This conclusion was based on the different training parameters explored during the early research. Some parameters such as image resolution play a vital role in the final result of the model and the loss calculated after each epoch. It is recommended to use the highest resolution the machines memory can afford during training and testing. Parameters which affect model training are:

- Image resolution - higher the better.
- Epochs: 200-500 is enough. Further training on same dataset will lead to over fitting.
- Number of image sequences that used for training and validation - more the better.

VI. CONCLUSION

In this study, we have trained RCNN based visual odometry DL model to track the trajectory of moving camera based on the RGB image sequence from that single camera. This Deep-MVO model is trained and tested by the KITTI dataset. The proposed model test result has been compared with two different approaches' results, and it showed lower rotational error in both cases.

The main component is the trained model comprising of the RCNN network updating the weights of the model.

This can be further trained and improved over time. This model will essentially act as a pipeline for end-to-end visual odometry.

The results are good as the trajectory is well traced but the start direction at the origin is flawed.

This will be improved with further training and by improving the parameters of the model itself.

This paper aims towards consistent results rather than accurate results so that the model can be used in unknown environments. Models such as these will be used more in space where visual sensors are highly reliable. With transfer learning, deep-MVO can be also used to track the accurate distance travelled in the trajectory and any other applications that can be derived from this process.

With huger datasets appearing online regularly, model such as these have a huge learning opportunity. Then they can be deployed on autonomous robots, self-driving cars or track the robot trajectory in locations where GPS signal is bad.

REFERENCES

[1] D. Wofk, F. Ma, T. -J. Yang, S. Karaman and V. Sze, "FastDepth: Fast Monocular Depth Estimation on Embedded Systems," 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 6101-6108, doi: 10.1109/ICRA.2019.8794182.

[2] S. Saleh and S. A. Khwandah, 'Traffic Signs Recognition and Distance Estimation using a Monocular Camera', 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 407–418, 2019

[3] Y. Cheng, M.W. Maimone and L. Matthies, "Visual Odometry on the Mars Exploration Rovers – A Tool to Ensure Accurate Driving and Science Imaging," IEEE Robotics & Automation Magazine, Jun 2006, pp. 54-62

[4] A. Geiger, J. Ziegler and C. Stiller, "StereoScan: Dense 3d reconstruction in real-time," 2011 IEEE Intelligent Vehicles Symposium (IV), 2011, pp. 963-968, doi: 10.1109/IVS.2011.5940405.

[5] R. Giubilato, M. Pertile and S. Debei, "A comparison of monocular and stereo visual FastSLAM implementations," 2016 IEEE Metrology for Aerospace (MetroAeroSpace), 2016, pp. 227-232, doi: 10.1109/MetroAeroSpace.2016.7573217.

[6] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," in IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255-1262, Oct. 2017, doi: 10.1109/TRO.2017.2705103.

[7] J. Sturm, N. Engelhard, F. Endres, W. Burgard and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 573-580, doi: 10.1109/IROS.2012.6385773.

[8] S. Shadi, S. Hadi, M. A. Nazari, and W. Hardt, 'Outdoor Navigation for Visually Impaired based on Deep Learning', 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 397–406, 2019.

[9] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, "Method: DeepVVO [DeepVVO]," Nov 2020, [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry_detail.php?&result=2783187176c31e31fff74b4f6b3fc13eb65c6037 [Accessed : 01-Aug-2021].

[10] R. Zhu, M. Yang, W. Liu, R. Song, B. Yan and Z. Xiao "DeepAVO: Efficient Pose Refining with Feature Distilling for Deep Visual Odometry," arXiv:2105.09899, 2021

[11] M. H. Mirabdollah and B. Mertsching, "Fast techniques for monocular visual odometry," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 297–307.

[12] I. Cvišić, I. Marković, and I. Petrović, "Recalibrating the KITTI Dataset Camera Setup for Improved Odometry Accuracy," arXiv:2109.03462, Sep 2021.

[13] A. Geiger, P. Lenz, C. Stiller, R. Urtasun "The KITTI Vision Benchmark Suite," Cvlibs.net. [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry.php [Accessed: 07-Sep-2021].

[14] S. Wang, R. Clark, H. Wen and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks," 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 2043-2050, doi: 10.1109/ICRA.2017.7989236.

Toward CNNs Visualization for Estimation Depth from Monocular Camera

Shweta Dumbre
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
shweta.dumbre@s2018.tu-chemnitz.de

Shadi Saleh
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
shadi.saleh@informatik.tu-chemnitz.de

Batbayar Battseren
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
batbayar.battseren@informatik.tu-chemnitz.de

Wolfram Hardt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
wolfram.hardt@informatik.tu-chemnitz.de

***Abstract*—The aim of a depth estimation is to achieve a representation of a scene's spatial structure and retrieve the three-dimensional shape and appearance in imagery representation. For depth predictions, we use an architecture that combines a transformer encoder and a convolutional decoder. We adopt dense vision transformers, an architecture which embraces vision transformers as a backbone also identified as encoder for dense prediction tasks instead of fully convolutional networks. We compile features named as tokens from different stages in the vision transformer to represent images at different resolutions and gradually combine them into a convolutionary decoder to establish full resolution predictions. The transformer as encoder has a global receptive field at all stages and provides constant and relatively high resolution representations. These characteristics provide fine-grained information and more consistent predictions in relation to fully-convolutionary networks. Using the KITTI and NYUv2 datasets, we show that the proposed technique outperforms state-of-the-art techniques by a wide margin. For estimates of monocular depth, the relative output of this architecture yields better depth predictions compared with state-of-the-art monocular depth estimation with a fully convolutional network.**

Keywords— Monocular depth estimation, Vision Transformer, fully-convolutional network, global receptive field component

I. INTRODUCTION

Image depth estimation is an essential role in computer vision, which can be commonly used in simultaneous localization and mapping (SLAM) navigation, object detection, image classification and semantic segmentation etc. [1]. It remains difficult and crucial to measure the distance from a camera to object. Applications such as robotics and autonomous driving, depth is a crucial requirement for perceiving, navigating and preparing for various tasks. Depth is also crucial in applications such as Road signs are an important part of the transportation system infrastructure. They provide road users with critical driving information. This, in turn, necessitates them adapting their driving habits and rules of the road [2]. Outdoor obstacles are usually unknown to visually impaired and blind people. In order to avoid collision hazards, they require guidance [3]. Depth is derived from two common methodologies in computer vision. That is, monocular images from monocular camera and stereo images from stereo camera by using epipolar geometry. The depth from stereo camera, Stereopses (stereo vision) are one of the most well-known strategy's in which depths are calculated with triangulation method using left and right images. The Detect-Describe-Match pipeline is typically used for stereo or multiview depth estimates that requires triangulation. When the scene is shot from a different perspective or the lighting varies between photographs, matching becomes incredibly challenging [4].

Fig. 1. RGB image and corresponding Depth image [1]

The Monocular Depth Estimation is the task of estimating scene depth using a single image, as shown in Fig.1. In order to obtain monocular depth map, it utilizes an image or video sequences. The main restriction of monocular depth estimation, is that the depth intrinsic properties are vanished during the scene's projection into the image plane. The advantage of this approach is that the operations required to process one image rather than two images such as left and right images are comparatively small. There are some benefits to estimating a single image depth, as depth estimation from multiple observations achieves impressive progress, it directly leads to depth estimation with a single image since it requests less expensive and geometric constraint. In certain areas, depth information is important such as perception, 3D modeling, robotics and autonomous movement [4]. The depth estimation from stereo images had significantly increased over the years. But in comparison to monocular depth estimates, stereo depth estimation needs more time and data.

Traditional monocular depth estimation was based on monocular cues. These monocular cues were, changes in texture and gradients, occlusion, object sizes and defocus. Because of these cues, humans tend to have astonishing capability to determine the depth from a single image. The local cues namely are haze (the atmospheric light scattering), defocus (blurring objects which are not focused) and occlusion (overlapped objects). So estimating depth only depends on local image properties [5].

The depth information of the image can directly be obtained by depth sensors, such as RGB-D cameras and LIDAR. RGB-D camera can directly capture the RGB pixel depth map, but they struggle with the reduced illumination sensitivity and measurement range. Though LIDAR is commonly used for depth measurement in the automated vehicle industry, it can only produce an incomplete 3D map. Furthermore, these large size and maximum energy consumption sensors (RGB-D cameras and LIDAR) affect applications including small robotics, such as drones. So as a replacement, the use of monocular camera began to emerge. The estimation of a dense depth map from a single image has gained much attention because of its low costs, the small scale, and the broad application of monocular cameras, and it has been recently researched in depth, based on deep learning [6]. Depth estimation's major goal is to extract a representation of the environment's spatial structure and to restore the three - dimensional structure and visual aspects of objects in imaging. As a result, we use a Monocular Camera to perform real-time 3D perception of the scene [7]. Nowadays, convolutional neural networks have gathered more attention for extracting depth information from single images. CNNs use encoder and decoder architectures. However, the option of encoder architecture has a massive impact on the general model. Information lost in the encoder cannot be recovered in the decoder. The convolutional encoder down samples means reducing the spatial resolution of the input image gradually to extract depth features on too many scales. Down sampling of the input image creates some problems, such as the receptive field increasing, grouping low-level features into concrete high-level characteristics, and simultaneously guarantees the traceability of network memory and computer requirements. Also, the disadvantage of down sampling is in prediction of depth. In the deepest stages of the model, the characteristics and granularity, also called fine-grained information, in the input image are lost. It becomes hard and difficult for the decoder to retrieve this necessary information. However, for some tasks, such as image classification, feature resolution and fine-detailed information in the input image are unimportant. But for the depth map, these characteristics play an important role [8].

The project introduces the dense transformer (DPT). DPT is a dense prediction architecture that uses the transformer as the encoder's basic computational based on an encoder decoder design. The latest proposed vision transformer (ViT) is used as encoder in this architecture. In contrast to fully convolutional network, vision transformer performs down sampling operations after the computation of initial image embedding stage and maintains constant dimensionality through all the stages of the ViT. In all the stages of ViT, global receptive field is present. It combines the image like features from all these stages of encoder and combine them in the convolutional decoder to obtain corresponding depth. All the features extracted from the encoder stages of transformer, no important information is lost [8].

II. Research Questions and Objectives

A. *Research Challenges*

Depth is a key to accomplish multiple tasks such as perception, navigation and planning etc. The main goal or challenge of the project is to able to find an effective network to perceive depth. Also visualize and evaluate the predicted depth with already existing state-of-the-art networks. There are plenty models to achieve depth, for instance deep learning methods and Vision transformers along with convolution networks.

B. *Literature Survey*

Image depth estimation has been a part of research for many years. It is possible to avoid deadly accidents by detecting obstructions early and accurately estimating their distance. Currently available object detection, on the other hand, ignores debris and other object classifications that were not included in the training procedure. The driving area, on the other hand, is monitored and recognized using active sensors such as LiDAR and RADAR, which are both expensive [9] [10].

It considers images taken from various perspectives, temporal sequences or fixed frame, static scene, and shifting lighting. These techniques are typically only beneficial when more than one input image of the scene is significant. We will concentrate on works relating to monocular depth estimation, in which there is only one input image and no observations about the scene geometry, image classification and extraction of objects. Fully-convolutional network is considered to be prototypical architecture for dense prediction. Over the years, variations of this building framework have been recommended, but all current architectures use convolution and sub-sampling as fundamental elements to learn multi-scale representations that can influence large dataset [11]. There are few papers that we have studied for literature survey similar to our topic as shown below in the following subsections.

1) *Depth Prediction using Multiscale Deep Network*

This paper describes a new strategy that tackles the problem by combining two stacks of deep network: one that generates a basic entire image global prediction, and another that improves it locally. To assist quantify depth relations rather than scale, we use a scale-invariant error. This method produces state-of-the-art performance on both NYU depth and KITTI, and matches detailed depth bounds without the requirement for super pixilation, by exploiting raw dataset as enormous sources of training data. The limitation is it does not include more real world scenarios (3D) geometry information such as surface normal. It does not provide original input resolution depth maps. Depth detailed information is not established [12].

2) *Depth Estimation by Zero-shot Cross-dataset Transfer*

Most frequently Monocular depth estimation depends is proved to achieve better results on a big training set. Due to the difficulties to obtain dense ground-truth depth a variety of dataset with specific characteristics and biases have evolved. This model trains with multiple dataset even if their labels are mismatched. It proposes a new training objective that is resistant to changes in variety of depth and scale, recommended using systematic multi objective learning to merge data from several sources, and stress the necessity of pre-training encoders on supplementary tasks. The experiment consists of five different training dataset, including a latest, huge data source: 3D films. It employs zero-shot cross-dataset transfer, which means evaluation on dataset that were not observed during training. The results obtained by this method shows that combining data from several sources improves monocular depth estimation significantly. This solution also outperforms rival methods across a wide range of dataset, establishing a new standard for monocular depth estimation [13].

Rather than estimating depth of the reflector, the network determines depth depending on the content displayed on it. The examples such as paintings, photographs and mirrors are not depth predicted properly. Strong edges can cause depth discontinuities. In rare cases, thin structures can be overlooked, and relative depth arrangement between unconnected objects in the image may not succeed. Because of the low resolution input images, the imprecise ground truth in reality is towards the far range. So, the image may have fuzzy background regions [13].

3) Multi-Scale Local Planar for Monocular Depth

Convolution layer and spatial pooling layers degrade the spatial resolution of the output in encoder decoder techniques, and so, numerous approaches such as skip connections or multi-layer deconvolutional networks are used to restore the original image resolution for successful dense prediction. So the network design, uses innovative local planar guidance layers situated at many stages of the decoder for more effective guidance of densely encoded features to the appropriate depth predictions. The output from the encoder are then combined to forecast depth in full resolution. In two ways, this differs from multiscale network and picture pyramid techniques [14]. The limitation is that the depth predictions have frequent artifacts which is an undesirable result. The main reason of such situation is because of the lack of ground truth across the dataset. Here they have used KITTI dataset [14].

III. Methodology

A. Basic knowledge of Transformer

Transformers have been mainly proposed for machine translation, and in many applications NLP have become the latest solution. A transformer is a deep learning method that incorporates the attention mechanism to weight the significance of each element of the input data differently. Its primary applications are in natural language processing and computer vision.

Architecture such as Transformers has been the paradigm of choice for the processing of natural languages (NLP). The main strategy is to pre-train a large corpus of text and then fine-tune it to a smaller task-specific data set. Thanks to the computational efficiency and scalability of transformers, extraordinary models with more than 100 Billion parameters have been made able to be trained. There is also been a growing interest in combining convolutional neural networks (CNNs) with different types of self-attention, for example, by augmenting feature maps for image classification [1] or by further processing the output of a CNN with self-attention, for example, for object detection [15], video processing [16, 17].

B. Architecture

The dense vision transformer is introduced in this section. We adopt the general encoder-decoder structure that has previously worked well for dense prediction. We use vision transformers as the backbone, demonstrating how the encoder's representation may be effectively turned into dense predictions and providing insight for the strategy's success [8]. In Fig.2. depicts a high-level view of the entire architecture. The left part of the images shows, the input image is converted to tokens also named as features (orange) by extracting non-overlapping image patches and then linearly projecting by using the ResNet-50 feature extractor (DPT-Hybrid). A positional embedding is introduced to the image embedding, as well as a patch-independent readout token (red). After the tokens are fed through several levels of transformers, at numerous resolutions, the reconstruction of tokens from various phases into an image-like representation (green) takes place. The right part depicts, fusion modules together (purple) gradually fuse and up sample the representations. The centre depicts, with 1/s the spatial resolution of the input image, tokens are combined into feature maps with Reassembler operation [8].

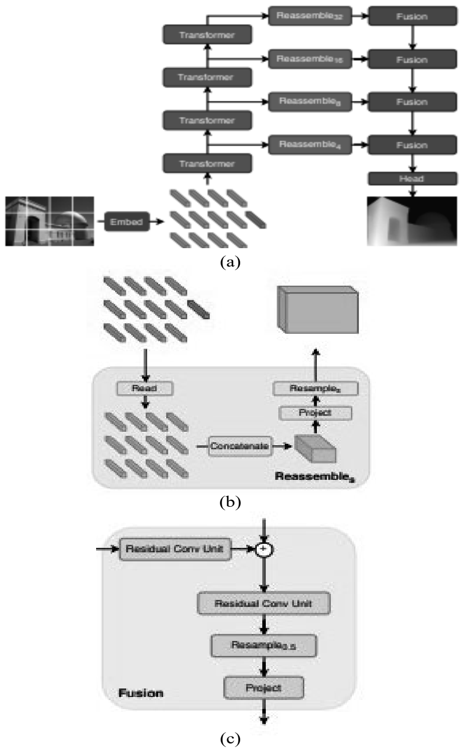


Fig. 2. Architecture Overview. Figure (a) shows the outline of the block diagram, Figure (b) shows Reassemble stages, Figure (c) shows Fusion stages [8]

C. Vision Transformer (ViT) Backbone (Encoder)

Initially, the image features are extricated from the input image. The role of "words" is played by image patches that are independently embedded in a feature space, or by deep features derived from the image. Throughout the report, the term "words" is known as "tokens". Transformers change the representation of a group of tokens by employing successive blocks of multi-headed self-attention (MHSA), which connect tokens to one another. For the application of depth estimation, the transformer takes account of all the tokens during all computations stages of transformer. Because tokens and image patches have a one-to-one correlation, the ViT encoder preserves the initial embedding's spatial resolution throughout all transformer stages. Furthermore, because every token may attend to and so influence every other token, MHSA is a fundamentally global active. In contrast to convolutional networks, the transformer after the initial embedding, possesses a global receptive field. The image patches from the input image are extracted using a ResNet50 variation of ViT, which employs the pixel features of the generated feature maps as tokens. Because transformers are set-to-set functions, they don't keep record of individual tokens spatial positions by default. To add this information to the representation, the learnable positions embedding is combined with image embedding's. Following NLP research, the ViT adds a specific token that isn't related to the input image but instead serves as useful to capture and distribute global information [8] [13].

$H \times W$ pixels is a set of $t^0 = t_0^0 \dots t_{N_p}^0, t_n^0 \mathbb{R}^D$ (1)

where (1), $N_p = \frac{HW}{p^2}$, t^0 =readout token, D = feature dimension of each token

The image embedding is computed using a ResNet50 which is followed by 12 transformer layers in ViT-Hybrid. For all trials, we utilize a patch size of p = 16. On contrary, to convolutional encoders, the ViT-Hybrid architecture extracts features 1/16 times the input resolution. It is very much larger than the normal convolutional encoders.

D. Hybrid Architecture

The patch embedding projection is applied to patches derived from a CNN feature map in this hybrid model. The input sequence can be produced by simply flattening the spatial dimensions of the feature map also called as tokens and projecting to the Transformer dimension in a specific scenario when the patches have a spatial size of 1x1[8] [13].

E. Convolutional Decoder

At varying resolutions, the decoder assembles the set of tokens into image-like feature representations and are gradually fused together to get the final dense prediction. To retrieve image-like representations from the output tokens of arbitrary layers of the transformer encoder, a simple three-stage Reassemble operation is performed [13].

$$Reassemble_s^{D1}(t) = (Resample_s \times Concatenate \times Read)(t) \quad (2)$$

where in (2), s denotes output size ratio and D1 denotes output feature dimension. Then the mapping begins Np + 1 tokens to a set of Np tokens that can be spatially concatenated into an image-like representation:

$$Read: \mathbb{R}^{(N_p+1\times D)} = \mathbb{R}^{(N_p\times D)} \quad (3)$$

The operation (3) is extremely responsible for handling the readout token. Then evaluate three different variants by using equation (3) for mapping.

$$Read_{ignore}(t) = t_{1,\dots\dots t_{N_p}} \quad (4)$$

The equation (4) ignores the readout token,

$$Read_{add}(t) = t_1 + t_0, \dots . t_{N_p} + t_0 \quad (5)$$

The equation (5) provides the information from the readout token to all other tokens by adding the representations,

$$Read_{proj}(t) = mlp(cat(t_1, t_0)), \dots . mlp\left(cat(N_p, t_0)\right) \quad (6)$$

The operation of (6) prior to actually projecting the representation to the original feature dimension D using a linear layer followed by a GELU non-linearity, transmits information to the other tokens by concatenating the readout to all other tokens.

Following a Read block, the Np tokens can be molded into an image-like representation by inserting each token in the image according to the position of the initial patch. Formally, we perform a spatial concatenation operation, which yields a feature map, as shown in (7),

$$Concatenate: \mathbb{R}^{N_p\times D} = \mathbb{R}^{\frac{H}{p}\times\frac{W}{p}\times D} \quad (7)$$

Finally pass the representation to the resampler that scales the representation as shown in Fig.3.

$$Resample: \mathbb{R}^{\frac{H}{p}\times\frac{W}{p}\times D} = \mathbb{R}^{\frac{H}{p}\times\frac{W}{p}\times D} \quad (8)$$

For ViT-Hybrid generates feature maps of 256 dimensions. Using a RefineNetbased feature fusion, combine the collected feature maps from successive stages and up sample the representation by a factor of two in each fusion stage. The resolution of the final representation is half that of the input image. To make the final prediction, connect a task-specific output head, as shown in Fig.3.

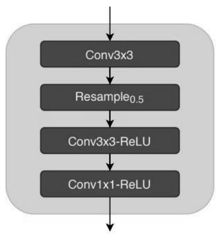


Fig. 3. Monocular depth estimator Head [8]

MIX-6, the network was also trained with an affine-invariant loss. The monocular depth estimation has been complied with this largest meta-dataset for the first time, as it contains 1.4 million images. Adam optimizer is used to combine multi objective optimization and set an encoder learning rate of 1e5 and 1e4 for the decoder weights. The

ImageNet weights, are used to pre-train the encoder, but the decoder is randomly initialized. An output head with three convolutional layers is used. After the first convolutional layer, the output head as shown in figure3.2 gradually reduces the feature dimension and up samples the predictions to the input resolution. Because of the negative impact of regression results, batch normalization is disabled. The images are scaled to 384 pixels on the longest side and trained on random 384-pixel square slices. The images are trained for 60 epochs with a batch size of 16, each epoch containing 72,000. Since, the batch size is not divisible by the size of the large dataset, so it takes subset of dataset to train for 60 epochs [8]

On the KITTI and NYUv2 dataset, fine-tuning of DPT-Hybrid is performed. So as to determine depth performance of DPT on smaller dataset such as KITTI and NYUv2. Because the network was trained with an affine-invariant loss, its predictions can be scaled up with enormous magnitude. Direct fine-tuning would thus be difficult, as the loss would be dominated by the overall mismatch in the size of the predictions to the ground truth.

As a result, robust alignment approach outlined in to match the initial network's predictions to each training sample. The resulting scales and shifts are then aggregated throughout the training set, and are applied to the depth predictions before delivering the result to the loss. The loss proposed by [12] is fine-tuned. Because KITTI only delivers sparse ground truth, we deactivate the gradient-matching loss [8] [13].

V. RESULTS

Our tests suggest that this design improves dense prediction tasks significantly, particularly when a large amount of training data is available. When compared to a state-of-the-art fully-convolutional network, we see a relative performance boost for monocular depth estimation.

The proposed method has been tested on the pre-trained DPT-Hybrid model on KITTI and NYUv2 dataset. Also have obtained the corresponding faithful depth predictions for both datasets. It provides stereo images, monocular images and 3D laser scans of outdoor scenes captured equipment mounted on the moving car. For KIITI, number of test images are 1000 and image resolution is 1216x352 pixels. The figure 5.1, shows KITTI RGB input image. As we can observe, the depth estimation from the proposed model or our model shows distinctive and fine-grained results as compared to other two models. The NYU-Depth V2 data set is comprised of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft Kinect. Total number of test images are 654 and image resolution is 640x480 pixels. The Fig. 4, illustrates KITTI result and Fig. 5, illustrates, NYU results.

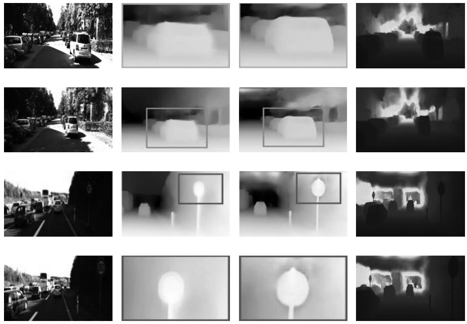
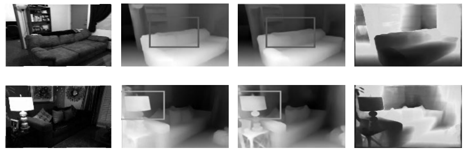


Fig. 4. Qualitative results on the KITTI test dataset (a) RGB image, (b) Yin et [18] (c) BTS [14] (d) Proposed depth image



(a) (b) (c) (d)

Fig. 5. Qualitative results on the NYU test dataset (a) RGB image, (b) Yin et [18] (c) BTS [14] (d) Proposed depth image

VI. EVALUATION

We try to compare and evaluated the proposed best performing model to various state-of-the-art approaches, as shown in Table 1, for KITTI and Table 2, for NYUv2. Both table depicts that, our proposed model outperforms the other two model based on the error metrics. The smaller fine-tuned model outperforms the competition by a significant margin. This demonstrates that our model's outstanding performance is related to enhanced network capacity and also to the presented training method. We have compared our model with "From Big to Small: Multi-Scale Local Planar Guidance [14]." The difference between proposed model and the other model is that, the proposed model has encoder as transformer and convolutional decoder with does not down samples the image resolution and provides faithful depth predictions.

The other model has convolutional encoder and decoder architecture with skip connections and decoder with multilayer deconvolutional layers to up sample the input images resolution, but still is not able to achieve high resolution depth predictions. We also compare our model with "Enforcing geometric constraints of virtual normal for depth prediction' [18]." The paper demonstrates necessity of high order 3D geometric restrictions for depth prediction. Unlike proposed model, the model does not work on large scale dataset. Depth predictions does not tent to provide faithful results. In depth image, the presence of some artifacts is also visible. For evaluation, we utilized error metrics as employed below with equations,

$$RMSE = \frac{\sqrt{\sum_{i=1}^{N} = (x_i - \hat{x}_i)^2}}{N} \tag{9}$$

$$MSE = \frac{\sum_{i=1}^{N}=(x_i-\hat{x}_i)^2}{N} \quad (10)$$

TABLE I. COMPARISON BETWEEN PROPOSED MODEL AND TWO STATE OF THE ART MODEL FOR KITTI

Method (KITTI)	MSE	RMSE
DPT-Hybrid (proposed)	0.010	0.101
Multi-Scale Local Planar Guidance for Monocular depth estimation [14]	7.595	2.756
Enforcing geometric constraints of virtual normal for depth prediction [18]	10.614	3.258

TABLE II. COMPARISON BETWEEN PROPOSED MODEL AND TWO STATE OF THE ART MODEL FOR NYU

Method (NYU)	MSE	RMSE
DPT-Hybrid (proposed)	0.038	0.061
Multi-Scale Local Planar Guidance for Monocular depth estimation [14]	0.153	0.392
Enforcing geometric constraints of virtual normal for depth prediction [18]	0.173	0.416

VII. CONCLUSION

Due to the introduction of Deep Learning algorithms in recent years, monocular depth estimation has progressed rapidly. This resulted in the development of transformers, which were originally used to perform image classification and object detection tasks but are now also used to perform tasks such as depth estimation with advanced transformers called as Vision transformers.

The dense prediction transformer or DPT, uses visual transformers to perform dense prediction tasks. When trained on large-scale dataset, DPT model reaches its maximum potential. When we compared to fully-convolutional architectures our model on monocular depth estimation provides more fine grained and globally coherent predictions. When pre-training on large dataset, this network works remarkably effectively. As a result, Vision Transformer meets or outperforms the state-of-the-art networks on tasks such as depth predictions while requiring less computational resources to train. Vision transformers success originate not just from their scalability, but also from enormous data self-supervised pre-training.

VIII. FUTURE SCOPE

While the results and evaluations are very much promising. But in future work, further scaling of ViT would likely lead to a better performance in depth predictions. Training of large scale data takes maximum time consumption and requires massive GPU support. These both factors can be improved in the future. Also ViT can be used for other tasks such as semantic segmentation.

REFERENCES

[1] ChaoQiang Zhao, QiYu Sun, ChongZhen Zhang, Yang Tang, and Feng Qian. Monocular depth estimation based on deep learning: An overview. Science China Technological Sciences, 63(9):1612–1627, Jun 2020.

[2] S. Saleh and S. A. Khwandah, ‘Traffic Signs Recognition and Distance Estimation using a Monocular Camera’, 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 407–418, 2019.

[3] S. Shadi, S. Hadi, M. A. Nazari, and W. Hardt, ‘Outdoor Navigation for Visually Impaired based on Deep Learning’, 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 397–406, 2019

[4] Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.

[5] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video, 2017.

[6] Chen Fu, Christoph Mertz, and John M. Dolan. Lidar and monocular camera fusion: On-road depth completion for autonomous driving. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pages 273–278, 2019.

[7] S. Saleh, S. Manoharan, and W. Hardt, “Real-time 3D Perception of Scene with Monocular Camera,” Embed. Selforganising Syst., vol. 7, no. 2, pp. 4–7, Sep. 2020.

[8] Ren´e Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction, 2021.

[9] S. Saleh, S. Manoharan, J. Nine and W. Hardt, "Towards Robust Perception Depth Information For Collision Avoidance," 2020 IEEE Congreso Bienal de Argentina (ARGENCON), 2020, pp. 1-4, doi: 10.1109/ARGENCON49523.2020.9505483.

[10] S. Saleh, S. Manoharan, J. Nine and W. Hardt, "Towards Robust Perception Depth Information For Collision Avoidance," 2020 IEEE Congreso Bienal de Argentina (ARGENCON), 2020, pp. 1-1, doi: 10.1109/ARGENCON49523.2020.9505566.

[11] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, and Nassir Navab. Deeper depth prediction with fully convolutional residual networks. In 2016 Fourth International Conference on 3D Vision (3DV), pages 239–248, 2016.

[12] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network, 2014.

[13] Ren´e Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer, 2020.

[14] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation, 2020.

[15] Han Hu, Jiayuan Gu, Zheng Zhang, Jifeng Dai, and Yichen Wei. Relation networks for object detection, 2018.

[16] Chen Sun, Austin Myers, Carl Vondrick, Kevin Murphy, and Cordelia Schmid. Videobert: A joint model for video and language representation learning, 2019.

[17] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Nonlocal neural networks, 2018.

[18] Wei Yin, Yifan Liu, Chunhua Shen, and Youliang Yan. Enforcing geometric constraints of virtual normal for depth prediction, 2019.

Emotion Identification Research for Neural Network Training

Sofia Kharyutkina
Department of Computer Engineering
Novosibirsk State Technical University
Novosibirsk, Russia
sonya.kharyutkina@gmail.com

***Abstract*—The paper sets the task: to study and select human emotions as feedback for training neural networks with the smallest error. To achieve this goal, the scientific literature was studied. In the scientific literature on the topic of emotion recognition, a program has been found to determine the basic emotions of a person. Based on the program, experiments were set up and conducted to study emotions. Emotions were chosen for the study: angry, disgust, scared, happy, sad, surprised, neutral. During the experiments, human emotions were determined, recognized by the program with the greatest accuracy. The average values of successful or unsuccessful recognition were calculated, and the similarity of emotions was analyzed. Thus, the relevance of the topic is given in the work, the program is described, as well as the libraries that it uses, an experiment is set and conducted, conclusions about the results obtained are given, as well as further plans in research.**

***Keywords*— emotions, neural network, program, error rate**

I. Introduction

Emotions are usually called a special kind of mental processes that express an individual's reactions to the effects of internal and external stimuli, having a pronounced subjective coloring in the form of direct experiences. Emotions cover all types of sensitivity and are an important factor in the regulation of vital activity. People are able to express a wide range of emotions, but there is no unambiguous definition and method of measuring emotion, which leads to different approaches to categorizing emotional data [1].

In recent years, the direction of affective computing has become particularly relevant – new technologies that provide human-machine interaction through the analysis of emotions [1]. This is an area of research aimed at developing intelligent systems capable of processing, recognizing and interpreting a person's emotional state, and subsequently adapting their behavior appropriately.

The classification of emotions in facial recognition (FER) is an important task of computer vision. It is aimed at detecting people's emotional states and intentions, and can be applied in many areas, such as social robots, medical care and other human-computer interactions. Emotions can usually be divided into seven categories: anger, disgust, fear, happiness, sadness, surprise and neutral mood [2].

Emotions, as a kind of physiological and psychological state of a person, play a vital role in a person's daily life [3]. Emotion recognition is important for: the work of call centers; the treatment of mental disorders; the treatment of Parkinson's and Alzheimer's diseases; deception recognition, online learning and the development of applications for games; stress analysis and the synthesis of emotions in a more realistic environment [4]. Taking out features is an important step in the analysis of expressions, which contributes to the rapid and accurate recognition of expressions, that is, happiness, surprise and disgust, sadness, anger and fear are expressions of faces. Facial expressions are most often used to interpret human emotions. Two categories contain a range of different emotions: positive emotions and non-positive emotions [5].

Emotions have no clear boundaries, so the task was set in the work: to select several emotions that would not overlap with each other as much as possible in terms of recognition and use them to train a neural network.

II. Description of the Laboratory Stand

As mentioned above, a program was taken to solve the task, the code of which can be found at the link https://github.com/omar178/Emotion-recognition.git. The program is written in python. In real time, it has the ability to recognize emotions: anger, disgust, fright, happiness, sadness, surprise, neutral emotion. It calculates the probability of what kind of emotion the person in the frame is experiencing. The program is based on the Facial Expression Recognition 2013 (FER-2013) database. The database stores 35,887 images with a size of 48x48 pixels, divided into 7 classes (anger, disgust, fear, joy, sadness, surprise, neutral).

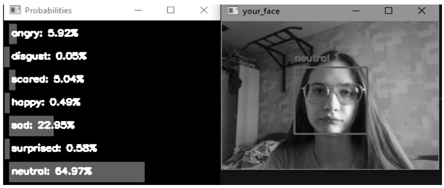


Fig. 1. An example of how the program works

The program selected for emotion research uses the following libraries: opencv_python, Keras, pandas, numpy, imutils, scikit_learn.

III. Description of the Experiment

The work is aimed at selecting emotions that can be recognized and used for further control of neural network learning. Let's conduct an experiment to see how accurately each of the seven emotions can be recognized by the program, that is, calculate the probability of which emotion it is. 10 experiments were conducted for each emotion. The following are the minimum and maximum probability values, as well as the calculated average value obtained as a result of the experiments.

The first emotion is anger. For the angry emotion, the values range from 40 to 70%. The average probability of angry was 56.2%.

The second emotion is disgust. For disgust, the values range from 60 to 95%. The average value is 78.7%.

The third emotion is scared. The values range from 20 to 50%. The average probability of scared was 35.1%.

The fourth emotion is happy. The values range from 80 to 99%. The average probability of happy was 87.6%.

The fifth emotion is sad. The values range from 20 to 40%. The average sad probability was 30.3%.

The sixth emotion is surprised. The values range from 60 to 80%. The average value of the probability of surprised was 71.5%.

The seventh emotion is neutral. The values range from 70 to 90%. The average probability of neutral was 81%.

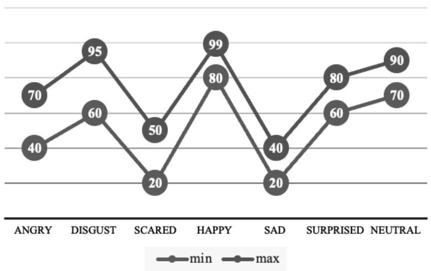


Fig. 2. A graph with the resulting ranges in percentages for each emotion

The graph shows the emotions being studied, as well as a range of values. The minimum values are highlighted in blue, and the maximum values are highlighted in red.

IV. Discussion of the results

Now we will summarize the data obtained in a single table. Table 1 shows the average values of the probabilities of the emotions considered.

TABLE I. THE RESULTS OBTAINED

Emotion	Average value, %
angry	56.2
disgust	78.7
scared	35.1
happy	87.6
sad	30.3
surprised	71.5
neutral	81

Thus, based on the data obtained, it can be seen that the program best determines the emotion happy, then disgust and neutral. The worst defines sad and scared. It can be concluded that using emotions as a reward and punishment, you can choose, for example, combinations: happy and angry; happy and disgust; happy and sad.

Conclusion

Three emotions recognized by the program with the smallest error are determined: happy, disgust, neutral. Due to the fact that each person is unique and expresses his emotions in his own way, in the future it is planned to increase the number of people tested, taking into account gender and race, to see how well the program will cope with recognizing the emotions of other people. And based on the results obtained, draw conclusions about the number of emotions used for further work.

References

[1] Ryumina E.V., Karpov A.A. / Analytical review of emotion recognition methods by human facial expression // Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics. – 2020. – P. 163-176.

[2] Deep reinforcement learning for robust emotion classification ib=n facial expression recognition / H. Li, H. Xu // Knowledge-Based Systems. – 2020. – N 204.

[3] R. J. Dolan / Emotion, cognition, and behavior // Science 298 (5596).– 2002.– P. 1191 – 1194.

[4] Investigation of multilingual and mixed-lingual emotion recognition using enhanced cues with data augmentation / S. Latitha, D. Gupta, M. Zakariah, Y. A. Alotaibi // Applied Acoustics. – 2020. – N 170.

[5] Emotion analysis using deep learning / D.S.L. Parvathi, N. Leelavathi, J.M.S.V. Ravikumar. B. Sujatha // 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).– 2020.– P. 593 – 598.

[6] Affective Computing.– https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_affective_computing.pdf. – P.90 -109.

[7] An overview and comparison of free Python libraries for data mining and big data analysis / I. Stancin, A. Jovic // Mipro. – 2019. P. 977–982.

[8] Nelli F. Python Data Analytics . – apress, 2018. – 576 c. – doi: https://doi.org/10.1007/978-1-4842-3913-1.

[9] Array programming with NumPy / Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, Travis E. Oliphant // Nature. – 2020. Vol 585. – P. 357–362. – doi: https://doi.org/10.1038/s41586-020-2649-2.

[10] Low Latency Image Processing of Transportation System Using Parallel Processing co-incident Multithreading (PPcM) / S. K. Shammi, S. Sultana, S. Islam, A. Chakrabarty // 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). – 2018.

[11] Scikit-image: image processing in Python / Stefan van derWalt, Johannes L. Schonberger, Juan Nunez-Iglesias, Francois Boulogne, Joshua D.Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu // PeerJ.– 2014. – doi: 10.7717/peerj.453.

[12] An overview of free software tools for general data mining / A. Jovic, K. Brkic, N. Bogunovic // Mipro. – 2014. – P. 1112–1117.

Training Spatial Ability Using Geogebra Interactive Dynamic Software

Suvd Erdenechimeg
School of Civil Engineering and Architecture
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
suvd.e@must.edu.mn

Batsuren Baasanjargal
School of Civil Engineering and Architecture
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
batsuren@must.edu.mn

Ganbat Danaa*
Open Education Center
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
ganbatda@must.edu.mn

***Abstract* - Descriptive geometry is one of the difficult subjects for students because students are expected to have the ability to visualize, read the 3D object from their 2D representations, draw a figure, and know the kinds of figures. According to some studies, if students have the weak-developed spatial ability, then they often face difficulty in understanding abstract concepts in the course. The present study conducted a short experiment and assessed the training experiment to improve spatial abilities using interactive dynamic software. The training focused on manipulation of dynamic geometric models and visualization of cross sectional views of 3D objects. The results indicated that training activities improved the subcomponents of spatial cross-section visualization ability. In addition, it was concluded that all students (either high or low spatial abilities) significantly improve spatial visualization ability measured with the Santa Barbara Solid Test after performing the proposed dynamic training.**

Keywords - Descriptive geometry; Santa Barbara Solid Test; Spatial visualization; Cross-section.

I. Introduction

Spatial ability is a major factor in human intellect. The ability to visualize objects and situations in one's mind, and to manipulate those images, is a cognitive skill vital to many career fields. Many studies have indicated the importance of spatial ability to the success of engineering students [1], [2]. Up to now, there is not any clear agreement on the subskills that this component is made up of. According to Lohman, spatial abilities are divided into spatial visualization, spatial orientation, and spatial relation [3] and these abilities help to understand graphical disciplines,especially descriptive geometry. Spatial visualization skills refer to the ability to encode and maintain spatial information in working memory while transforming it, which are valuable for several Science, Technology, Education and Mathematics (STEM) fields. For example, the ability to mentally represent sectional views of objects is correlated with one's capacity for spatial visualization [4], [5].

The ability to infer the external shape and internal features of sections of objects and structures plays an important role in many fields of science such as medical professionals must learn to infer the shapes of cross sections for understanding X-rays and magnetic resonance imagining (MRI) information whereas understanding the cross-sectional structure of materials and mechanisms is a fundamental skill in engineering [6]. Previous studies determined that the ability to infer a cross section of an object is positively correlated with spatial visualization ability [7], [8]. Spatial ability is one dimension of individual difference that can influence a learner's ability to extract information from dynamic, interactive animations [9].

Recent years, number of students who unsuccessful studied on descriptive geometry course have been increasing. For instance, in the last eight years at the Mongolian University of Science and Technology (MUST), over 50% of students learned unsuccessfully on the course of descriptive geometry [10], which is a fundamental principle of engineering and architecture, since it provides students with an intellectual capacity for spatial visualization. If the teaching and learning properly conducted , the Descriptive geometry develops the students' ability to imagine objects or designs in space and is not just concerned with the reading and interpretation of drawings. Without the ability to think in three dimensions and apply this into drawing, creativity and intelligence when designing new things are not feasible. Well-developed spatial visualization abilities are important conditions for all engineering studies [11]. Students often have difficulty in understanding abstract concepts in descriptive geometry. We predicted that the problem might be reasoned with students' low spatial ability.

II. Research Objective and Question

Considering above mentioned problem, the purpose of our study was to investigate the effects of using the interactive dynamic software GeoGebra on the spatial visualization skills of students who were studying Descriptive geometry. Thus, the following research question was addressed: Does using the interactive dynamic geometry program GeoGebra help students to develop their spatial visualization skills for both groups (high and low spatial abilities)?

III. Methodology

This experiment uses the Santa Barbara Solid Test (SBST) as a measure of performance before and after training of spatial visualization training. So, in the present study, a group pretest-posttest experimental design was utilized to examine the effect of interactive dynamic GeoGebra[1] software on the improvement of students' spatial visualization ability.

* Corresponding author

[1] GeoGebra (the name is made from the two words Geometry and Algebra) is an interactive geometry, algebra, statistics and calculus application,

A. Material

Santa Barbara Solid Test (SBST): The SBST is a 30 items multiple choice test in which participants are asked to solve spatial problems related to sectional views (Fig. 1b). The items consist of simple 10, joined 10 and embedded 10 geometric structures of solids while section planes that are orthogonal 15 (horizontal or vertical) and oblique 15 [13]. The test was designed based on cognitive theory that accounts for variability among students in the capacity to mentally form and manipulate visual images [14]. The test assesses the ability to identify the two-dimensional section view of a three-dimensional object.

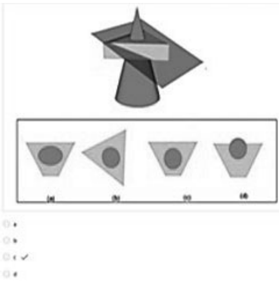


Fig. 1. The electronic version of the spatial ability test (SBST)

A total of 42 students who studied the Descriptive Geometry course at Mongolian University of Science and Technology participated in this study. However, 14 participants did not fully complete all instruments. Thus, only 28 participants (10 female, 18 male) were taken into consideration in the analysis.

In the pre-test result 13 of them had a high spatial abilities and other 15 of them had a low spatial abilities with SBST measurement. In this study, median (15.0) split was used to classify participant as having high or low spatial abilities[2]. This provides a pre-requisite way to categorize learners with different spatial abilities [12].

C. Procedure

The experiment was conducted on the eighth week of the spring semester of academic year of 2020-2021. At the beginning of the experiment, the students were administered the electronic version of SBST within MS Office 365 Form. After completing the test, they read the instructions of the training and the material available to perform it. The proposed GeoGebra activity is an effective way for improving spatial ability when students have to complete tasks.

Interactive dynamic learning materials prepared for Cross sectional topics of Descriptive geometry, the dynamic geometric model, was used as the web-based learning material. The training experiment was organized into three sessions with a total duration of 120 minutes (pre-test of 30 minutes for SBST, spatial training session of 75 minutes and post-test as finalizing session with 15 minutes). This training material was created using McKim's (1980) model of spatial visualization [15]. It consisted of three parts (seeing, imaging, and drawing) and each one of them contained several types of exercises. The examples of the training activities are shown in Fig 2.

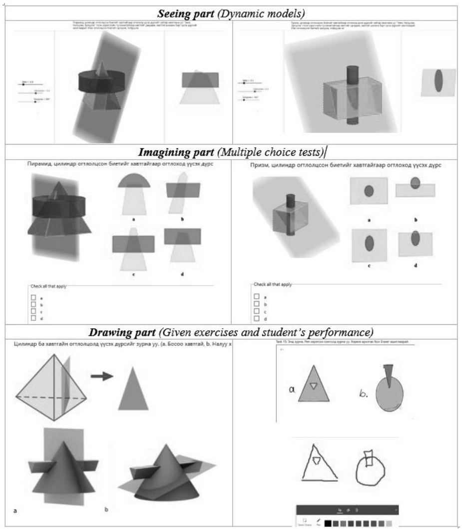


Fig. 2. Examples of each training activities

For the seeing part of students could freely manipulate (to rotate, to tilt and to move) and observe the given dynamic models. To the next part, students were asked to get a multiple-choices test and they could manipulate the interactive dynamic three-dimensional model in GeoGebra activity and check whether their performances of the imagining (visualizing) part corresponded to the 3D virtual models that they were visualizing. Whereas, in the drawing part sectional exercises should be completed without the help of any dynamic model and they were given the drawing tasks asked to free hand sketch and draw using the GeoGebra tools. This part included six of SBST structure based on drawing exercises and students needed to visualize how the section would look the sectional figures then to draw the sectional views of three geometric solids such as simple, joined and embedded. Section planes are given orthogonal (horizontal or vertical) and oblique. Students were asked to create the sectional views of these solids using a section plane as shown in the drawing part of Fig. 2. After the training, the students took the SBST as post-test again.

intended for learning and teaching mathematics and science from primary school to university level. GeoGebra is available on multiple platforms, with apps for desktops (Windows, macOS and Linux), tablets (Android, iPad and Windows) and Web. www.wikipedia.com

[2] Low spatial abilities SBST≤15 score.
High spatial abilities SBST>15 score.

IV. Results

A. Internal reliability

Considering that both pre and post SBSTs were satisfactory reliable with α=0.86 and α=0.81, respectively.

B. Descriptive statistics

Descriptive statistics were calculated for all variables such as SBST and its subscales and given in Table I. According to the result, students answered slightly more than half of the 30 items correctly (M =15.53, SD = 6.19) on SBST. For pre-score of low spatial group, it was (M =10.67, SD = 2.66), while for high spatial group was (M =21.15, SD = 3.76). Subsequently, we assessed whether there were differences in the gain values with respect to individual difference. To explore whether differences in learners' ability affect the improvement of spatial visualization ability, this study categorized the Pre-SBST scores (median) into high and low spatial abilities.

TABLE I. THE DESCRIPTIVE STATISTICS OF THE SBST AND ITS' SUBSCALES SCORES BEFORE TRAINING

Groups	*Low spatial abilities*			*High spatial abilities*		
Measures	*N*	*Mean*	*S. D*	*N*	*Mean*	*S. D*
Simple orthogonal	15	2.00	0.85	13	4.23	0.93
Simple oblique		1.67	0.72		2.77	1.09
Joined orthogonal		2.87	1.19		4.31	0.75
Joined oblique		0.67	0.72		2.38	1.50
Embedded orthogonal		2.53	1.06		4.77	0.44
Embedded oblique		0.93	0.88		2.69	1.32
Pre SBST		**10.67**	**2.66**		**21.15**	**3.76**

The students took the same SBST test again, after the training the data for the descriptive statistics obtained from the SBST and it is presented in Table II.

' SUBSCALES SCORES AFTER TRAINING

Groups	*Low spatial ability*			*High spatial ability*		
Measures	*N*	*Mean*	*S. D*	*N*	*Mean*	*S. D*
Simple orthogonal	15	3.27	1.03	13	4.08	0.76
Simple oblique		3.13	1.25		3.54	0.97
Joined orthogonal		3.87	0.99		4.62	0.65
Joined oblique		2.47	1.60		3.77	1.09
Embedded orthogonal		3.53	1.13		4.62	0.65
Embedded oblique		2.53	1.25		3.92	0.86
Post SBST		**18.80**	**4.78**		**24.54**	**3.18**

C. Effects of interactive dynamic training approach

According to the results in Table III, there is a significant difference in the students' spatial abilities in both low and high spatial group. For low spatial group, these differences were observed with respect not only to the test results taken as a whole, but also to each subscale of the test results ($p<.05$). For high spatial group, there were a significant difference of oblique of cutting plane across three types of structure. Based on these statistical outputs, for both groups, it is clear that interactive dynamic activities have positive effects on the students' acquisition of spatial abilities.

TABLE III. SBST PAIRED SAMPLE T-TEST ON PRETEST-POSTTEST

Pairs	*Groups*	*Low spatial ability*				*High spatial ability*			
	Measures	*t*	*df*	*p*	η^2	*t*	*df*	*p*	η^2
1	*Simple orthogonal*	5.10	14	**.000**	0.49	-.51	12	.613	0.01
2	*Simple oblique*	5.04		**.000**	0.49	2.37		**.035**	0.17
3	*Joined orthogonal*	3.09		**.008**	0.26	1.29		.219	0.06
4	*Joined oblique*	3.47		**.004**	0.31	3.32		**.006**	0.29
5	*Embedded orthogonal*	3.87		**.002**	0.36	-.69		.502	0.02
6	*Embedded oblique*	4.98		**.000**	0.48	3.80		**.002**	0.35
7	*Total SBST*	**6.83**		**.000**	**0.63**	**3.35**		**.006**	**0.29**

We went on to compute effect size for both groups. Then we calculated the eta squared (η^2) as:

$$\eta^2 = \frac{t^2}{t^2 + N - 1}$$

0.62 (very large effect) [16]. The detailed statistics for both spatial groups were interpreted below.

The eta squared statistics showed large effects for all of the subscales and were 0.63 for the total SBST of the low spatial ability. We also found the effect size for high spatial group large effects for 3 oblique subscales such as the simple oblique (0.17), the joined oblique (0.29), and the embedded oblique (0.35), moderate effect for except joined orthogonal (0.06), small effect for the embedded orthogonal (0.02). The guidelines, proposed by Cohen [16] for interpreting this value are: 0.01 small effects, 0.06 moderate effect and 0.14 large effects. Given our eta squared values for both groups, we can conclude that our treatment was indeed effective.

V. Discussion and Conclusion

The descriptive statistics show that the total mean of before training SBST for low and high groups was 15.53. It should be noted that although in present study, median (15.0) split was used to classify participant as having high (46.4% of the students) or low (53.6% of the students) spatial abilities, but 57.01% of all students could not perform over 60% of the test. In other words, over half of students failed the spatial test, indicating that the SBST was challenging for our students, especially for oblique of cutting plane items. It was explained more detailed following paragraph.

For the low spatial group, the highest mean performance was on the Joined orthogonal items (M=2.87); the lowest mean performance was on joined oblique items (M=0.67) on pre SBST, whereas, for high spatial ability group, the highest mean performance was on embedded orthogonal items (M=4.77); the lowest mean performance was on joined oblique items (M=2.38). For the high spatial ability group, the highest mean performances were on Joined and Embedded orthogonal items (M=4.62); the lowest mean performance was on simple oblique items (M =3.54) on post SBST, whereas, for low spatial ability group, the highest mean performance was on joined orthogonal items (M=3.87); the lowest mean performance was on joined oblique items (M=2.47). According to these outputs, with oblique plane's problems were consistently more difficult than orthogonal problems across three levels of geometric structure on both of the pre and post SBST, and we can see that easily from Table II and III. High spatial ability students consistently outperformed low spatial ability learners across pre and post-tests as well. Fig.3. demonstrates that the mean of the students' SBST scores increased in the post-test for both groups.

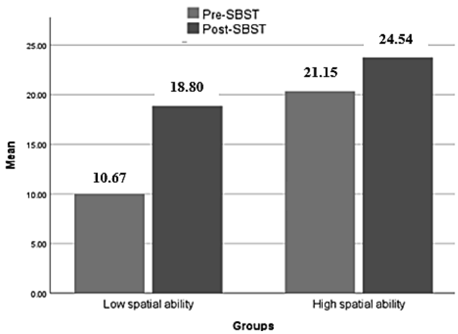


Fig. 3. Performance means by comparisons of Pre and Post SBST, by individual differences.

High spatial abilities could use their ability to compensate in an environment without explicit presentation of 3-D representation and dynamic visualization, but low spatial abilities could not [17]. Therefore, the ability as compensator hypothesized that low spatial ability learners should gain particular benefit from the interactive 3D virtual learning environment as they have difficulty to mentally construct their own visualization. The explicit presentation of 3D representations and dynamic visualizations may keep the need for using spatial processing schema to very minimum, thus reduces the extraneous cognitive load and fosters learning. Thus, this could help educators to facilitate individualized learning and teaching materials and support provided to learners should vary with their spatial abilities and task difficulty.

The present study shows that the interactive dynamic learning environment benefits to both of the low and high spatial ability learners. The interactive dynamic training provides students to learn cross sectional concept and to explore geometric relationships easily. Thus, this result shows training designed by interactive dynamic GeoGebra software has a positive effect on improvement for spatial visualization abilities of students. The significant positive effect of the interactive dynamic learning environment on the performance outcome has provided empirical evidence of the potential of dynamic GeoGebra software to support and enhance learning abilities of students in Descriptive Geometry.

VI. Limitation and Future Direction

The limitation of this study that only measured learning effectiveness for improvement of students' spatial visualization ability using interactive dynamic software for one group. Future works for Descriptive geometry teaching and learning, it is needed to include integration of the dynamic geometry software GeoGebra into the educational process and could include pre and post-test comparisons for two groups that include students' motivation and cognitive load.

VII. Acknowledgements

The authors acknowledge the support of students who participated in this study. Our deepest appreciation to Prof. Tudevdagva Uranchimeg for valuable guidance for improving our research work.

VIII. References

[1] D. I. Miller and F. D. Halpern, "Spatial Thinking in Physics : Longitudinal Impacts of 3-D Spatial Training," pp. 3465–3470, 2008.

[2] J. L. Mohler, "Examining the spatial ability phenomenon from the student's perspective," *Eng. Des. Graph. J.*, vol. 72, no. 3, pp. 1–15, 2008.

[3] D. F. Lohman, I. I. Dennis, and P. T. Eds, "Spatial ability and g," p. 2016, 1996.

[4] S. A. Sorby, "Developing 3-D Spatial Visualization Skills," *Engeneering Des. Graph. J.*, no. spatial visualisation, 1999, [Online]. Available: https://diggingdeeper.pbworks.com/f/Developing+Spatial+Skills.pdf.

[5] J. B. Carroll, *Human cognitive abilities-a survey of factor analytic studies*. Cambridge University Press, 1993.

[6] S. A. Sorby, "Assessment of a 'new and improved' course for the development of 3-D spatial skills," *Eng. Des. Graph. J.*, vol. 69, no. 3, pp. 6–13, 2005.

[7] C. A. Cohen, "The influence of spatial ability on the use of dynamic, interactive animation in a spatial problem-solving task," *AAAI Spring Symp. - Tech. Rep.*, vol. SS-05-06, no. January 2005, pp. 1–5, 2005.

[8] D. Bairaktarova and D. Williams, "Analysis of Blended and Multi-modal Instruction and its Effects on Spatial Visualization Ability.," *Eng. Des. …*, vol. 83, pp. 14–26, 2019, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=00462012&AN=141183820&h=%2F2wsXFvjJ%2B1OC6330XLaD7ho137lpoFsg%2BXd6WBgWGTK4%2FrhUDK9DfHky7Pdi9hNg7sfxQedvpXmx%2FEIDTG3NA%3D%3D&crl=c.

[9] C. A. Cohen, “The influence of spatial ability on the use of dynamic, interactive animation in a spatial problem-solving task,” *AAAI Spring Symp. - Tech. Rep.*, vol. SS-05-06, no. March, pp. 1–5, 2005.

[10] K. Damchaasuren and N. Chuluunbaatar, “Analysis of Engineering Drawing lecturers’ assessment result (academic year of 2016-2017),” in *The Today and Future of Engineering graphics education*, 2017, vol. 1, pp. 29–34.

[11] C. Leopold, “Geometry education for developing spatial visualisation abilities of engineering students GEOMETRY EDUCATION FOR DEVELOPING SPATIAL des métiers ’ (1751-1772) played an important role in this historical development . In the cur- common language for engin,” no. May, 2014.

[12] M. Kozhevnikov, M. Hegarty, and R. E. Mayer, “Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers,” *Cogn. Instr.*, vol. 20, no. 1, pp. 47–77, 2002, doi: 10.1207/S1532690XCI2001_3.

[13] C. A. Cohen and M. Hegarty, “Inferring cross sections of 3D objects: A new spatial thinking test,” *Learn. Individ. Differ.*, vol. 22, no. 6, pp. 868–874, 2012, doi: 10.1016/j.lindif.2012.05.007.

[14] T. J. Barrett and M. Hegarty, “Effects of interface and spatial ability on manipulation of virtual models in a STEM domain,” *Comput. Human Behav.*, vol. 65, 2016, doi: 10.1016/j.chb.2016.06.026.

[15] R. H McKim, *Experience in visual thinking*, 2nd ed. California: Monterey, California, 1980.

[16] J. Cohen, *Statistical Power Analysis for Statistical Power Analysis for the Behavioral Sciences*, Second. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers, 1988.

[17] E. A. L. Lee and K. W. Wong, “Learning with desktop virtual reality: Low spatial ability learners are more positively affected,” *Comput. Educ.*, vol. 79, pp. 49–58, 2014, doi: 10.1016/j.compedu.2014.07.010.

Implementation of the Online Professional Development Program for Mass University Teachers During COVID-19 Pandemic

Tserenchimed Purevsuren
Open Education Center
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
Tserenchimed.p@must.edu.mn

Tsooj Shambaljamts
dept. of Technocal Mechanics
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
tsooj@must.edu.mn

Ariunbolor Davaa
Open Education Center
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
ariunbolor@must.edu.mn

Ganbat Danaa
Open Education Center
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
ganbatda@must.edu.mn

Tumenbold Dashzeveg
dept. of Industrial design
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
tumenbold@must.edu.mn

Naidandorj Radnaa
Open Education Center
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
Naidan@must.edu.mn

***Abstract*—In order to enhance quality of online learning that surged after COVID-19 pandemic, the university teachers needs to be trained for effective online teaching and learning skills and ability to work efficiently in New Normal situation. The Open Education Center of Mongolian University of Science and Technology was implemented "Online Professional Development Program" during pandemic situation. In this study, we purposed to share our experience of implementing such online professional development program, used methodologies, and assessment result of the training program.**

Keywords—teacher development, digital transformation, pandemic, program assessment

I. INTRODUCTION

The COVID-19 pandemic impacted the education sector all over the world. In order to limit social distancing, a number of countries were closed the education institutions, therefore rapid paradigm shift from face-to-face to online learning was only choice for safe and sustainable education [1]. For Mongolian higher education alone, about 7 thousand teachers and 147 thousand students were engaged in online learning mode during pandemic. However everybody including teachers were not ready for such rapid shift [2]. The Mongolian government took several actions in education field during pandemic situation that included use of teacher training platform (www.esis.edu.mn), organizing Tele lectures (econtent.edu.mn) though these were targeted to the K-12 education [3]. While, there were no notable strategy and action against pandemic for higher education sector and higher education institutions were on their own during this rapid face-to-face to online learning shift [4].

The universities were utilized a different types of online learning format during pandemic. At early stage, majority of higher education institutions were relying on asynchronous online learning format by utilizing previously developed online contents, courses on the google classroom, Moodle, YouTube channels and respective learning management system (LMS). Solely use of asynchronous online learning based on existing educational digital resource, which is limited, was not sufficient enough. Thus, later on the universities were started to adapt synchronous format of online learning by utilizing online communication tools such as Zoom, Microsoft Teams, and Facebook. Currently, higher education institutions were mostly using hybrid format of asynchronous and synchronous leaning [4] for online teaching and learning process. However, the degree and quality of online learning is varied between each institutions.

A number of international studies were investigated the good practice for online learning in Covid19 situation and problems that universities are facing [5-14]. Lack of online learning infrastructure and digital contents, limited guidance for student and teachers, insufficient institutional readiness, and lack of teacher's knowledge to conduct online learning were the main problems [5-14]. It has been indicated that the most challenging issues were related to the university teachers such as a negative attitude towards the change [11], need of improving teacher's online teaching and learning practice [6, 7], and lack of digital skills [10, 12]. Thus, the first step to overcome pandemic situation successfully and to overcome current rapid shift is to strengthen existing professional development training by doing online and train the teachers as soon as possible.

The Open Education Center (OEC) of Mongolian University of Science and Technology (MUST) had developed "Online Professional Development Program (OPDP)" during Covid-19 pandemic and implemented nationwide training for university teachers. The purposes of current study were (1) to introduce the instructional design fundamentals of the developed program, (2) to share our experience of implementing the program, and (3) to demonstrate the assessment result of the training program that involved mass university teachers at MUST.

II. Materials & Methods

A. Planning and instuctional design of the OPDP

In order to successfully implement the OPDP, the OEC took one year for planning and preparation. The subsequent preparation works were determination of the online training program concept, driving purposed learning outcome, preparation or training of master teachers, choosing appropriate learning methodology as well as delivery method for online training, development of the digital contents, and selection of the online learning infrastructure or platform. Totally four team of 41 specialist and teachers were participated in the OPDP preparation and planning phase.

For preparation of master teachers who purposed to deliver overall OPDP, OEC strategically sent OPDP team members and young professors to the domestic and international online training programs, webinars, conferences; which related accelerating online learning and responded to the COVID-19 situation; such J-WEL Connection 2020 from Abdul Latif Jameel World Education Lab of MIT (https://jwel.mit.edu/), online trainings and webinars organized by International Centre for Higher Education Innovation under the auspices of UNESCO (http://en.ichei.org/), and a number programs delivered by European distance and E-learning network (http://www.eden-online.org/). The purpose of the master teachers at this stage were to gain knowledge about online learning fundamentals, to identify necessary contents that needed to be considered in the OPDP, to learn from international education practices that responded for COVID-19 pandemic, to localize online teaching and learning methods as well as recent educational technologies.

The next step was determination of desired learning outcome of OPDP. Domestic need analysis was performed within the teachers working at higher education sector of Mongolia. In addition, international studies published after COVID-19 were reviewed to identify necessary trends and issues needed to be considered in OPDP concept and its learning outcome. Following issues were identified:

- OPDP must provide digital skills to the teachers and enhance teacher readiness for online learning [12, 15],
- Purposed training program must be stimulation to develop self-learning community [16, 17] ,
- The program needs to provide fundamental theory behind online teaching and learning methodology [6, 18, 19],
- Purposed OPDP must be benchmark or best example of the online learning practice to the teachers [16],
- The concept of instructional design and its theoretical and practical knowledge must be provided during professional development training [6, 18, 19],
- The program must share theory and practice of online and distance learning to the participants [6, 18, 19],
- The teacher leadership and its importance in online learning should be discussed [10],
- To decrease negative attitude of teachers against the rapid change for online learning [11],
- To consider that the teachers have not much to self-learning time [5],
- The OPDP must be leverage of digital transformation at campus to accelerate current education paradigm shift.

The next step of planning the professional development online program is to identify key topics and contents that need to be delivered during training along with optimum delivery methods. In program, ten modules of education contents that can be delivered during two weeks were prepared based on domestic needs and international trends.

Week 1:

1. Module 1: Paradigms shift and education technology at 21st century
2. Module 2: Instructional design of the teaching and learning process
3. Module 3: New trends and direction on teaching and learning process
4. Module 4: Educational development in future
5. Module 5: Active learning in online learning

Week 2:

6. Module 6: Massive Open Online Course (MOOC)
7. Module 7: Open Educational Resource (OER)
8. Module 8: Use of the online learning platform MS Office 365 (MS- teams)
9. Module 9: Digital content development for educational purpose and its successful use in learning process
10. Module 10: Contribution of creating self-motivated community to the development of Open Education.

The delivery method of OPDP was designed as 2 week of online webinar trainings combined with professional development MOOC that prepared according to above mentioned topics (Fig. 1). The webinar type of training is generally a replication of traditional lecture. Thus, it is most likely being teacher-oriented information transferring activity rather than student-centered active learning process with learner engagement [20]. Therefore, in order to increase learner engagement to the program [21], active learning methodologies including team-oriented competitive learning, project-based learning, flipped classroom, and peer assessment methods are used during webinar training. The Microsoft Teams program was used as a webinar training tool. The webinar training was not enough for expected learning outcome of the OPDP and thus MOOC was developed using www.edx.edu.mn platform. The MOOC was expected to be a best source for self-directed learning as well as used as a tool for flipped classroom technique. The MOOC consisted of video lectures, reading materials, audio podcasts, self-checking test, and additional resource related to each modules. It can also act as an important tool for sustainable teacher development platform after delivery of the OPDP. Use of MOOC alone may lead increased learner dropout rate [22]. Therefore, combination of live webinar and standalone MOOC was best delivery strategy for OPDP.

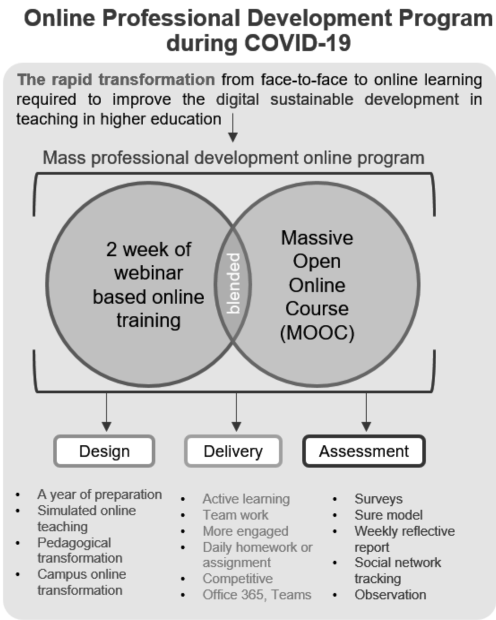


Fig. 1. Example of a figure caption. (*figure caption*)

B. Iimplementation and delivery of the OPDP

The pilot test of OPDP implementation was conducted by OEC in cooperation with Mongolian Institution of Teacher's Professional Development (ITPD) during 2020.12.7 to 2020.12.18 for 73 teachers from different Universities and Colleges. The purpose of pilot test was to investigate the possibility of mass further mass online training, lack and mistakes in prepared contents or topics, to encourage OPDP delivery team, and to check the reliability of online learning infrastructure and software. Totally 18 individuals including technician, professors, and master teachers were organized the pilot test.

C. Assessment and evaluation of the OPDP

There are two kind of program evaluation techniques were used for assessing teacher's satisfaction, successful program devilry process, and positive and negative post-training impact. First, teacher's satisfaction for quality of online training program was evaluated using survey collected during 2 week of webinar trainings. MS form program was utilized to collect individual surveys. The survey consisted of 6 questions for assessing quality training materials, digital contents and methodology; 7 questions for assessing training management and digital infrastructure; 3 questions for assessing participant active engagement; and 5 questions for assessing participant self-motivation. For each question, the maximum power was 5 point. In order to quantitatively evaluate the OPDP success, the Structure Oriented Evaluation Model for E-Learning (the SURE model) [23, 24] was utilized for analyzing the survey results. Finally qualitative research approach was used to evaluate the post-training impact after OPDP delivery on the online learning at MUST. In which, weekly report of the online learning was collected from every single teachers at campus. The thematic analysis technique was used to identify key challenges, problems in online learning. Moreover, teacher's social activities including posts, shares, and comments related online learning at social network (Facebook group of MUST teachers) was monitored after OPDP implementation.

III. RESULTS

Totally 784 professors (76.6% female, 23.4% male) were participated to the OPDP. 84% of them were from Ulaanbaatar, the capital, and 6% were from countryside campuses. 51-64% of overall participants were actively engaged daily events. Also, 69% of registered participants were engaged to the developed MOOC at www.edx.edu.mn. 1.8% of registered participants were inactive trainees. 61.6% of participants to the OPDP were master, 37% Ph.D., and 1.4% of them were Science Doctor.

The result of teacher's satisfaction survey was shown in the fig. 2. The result showed that teacher's satisfaction scores were ranged between 4.3±0.3 to 4.6±0.7 for quality of training materials and methods, 4.0±1.0 to 4.7±0.7 for management and digital infrastructure, 4.3±0.7 to 4.6±0.6 learner engagement, and 4.0±0.8 to 4.5±0.7 for participant self-motivation, respectively. The SURE model assessment coefficient (K_{sure}) for overall OPDP was average 0.95 (coefficient of one indicates 100%). It was 0.96 for quality of training materials and methods, 0.98 for management and digital infrastructure, 0.94 for learner engagement, and 0.92 for participant self-motivation, respectively. The lowest value of SURE model assessment coefficients were 0.81 for internet connection and stability, 0.81 for knowledge acquirement from delivered contents, 0.84 for enhancement of digital skills, and 0.84 for usability of the software and platform, respectively.

The thematic analysis results of weekly online learning report, which started collecting after program implementation, was shown in the Table I. For teacher, three main issues observed were the lack of equipment to conduct online learning, increased workload in online learning in comparison with traditional face-to-face learning, and health problems related long time working online. For student, the most problems were related to bad internet connection and high cost of data usage to connect MS Teams that we adapted as an online learning tool after OPDP (Table I). Due to high cost and bad internet connection (especially for student from countryside), student engagement and motivation to join courses were decreased dramatically. Students were mostly using mobile devices to online learning. Thus it was reported that importance of practical courses such as seminar and laboratory are decreased and more likely being similar to lectures. For management level, two main challenges were observed. The time overlap of online courses and other meetings such as department meeting, management team meetings, human resource training, and thesis defense was the one of the problem for teacher's time management. Also, teachers were complaining about solely use of old LMS and neglecting the use of newly adopted MS-team for online learning quality assurance and teachers workload assessment mechanism

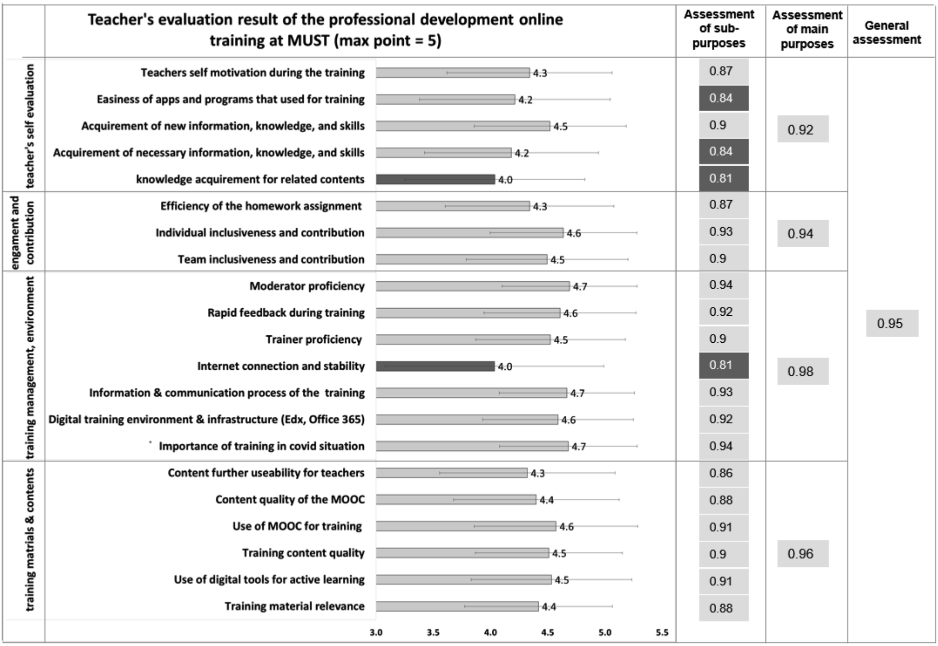


Fig. 2. The survey result for OPDP assessment (max point is 5 for each question). The SURE model coefficients, K_{sure} [23, 24], are listed at right column.

The result of monitoring teacher's social activities including posts, shares, and comments related online learning at social network (Facebook group of MUST teachers) was summarized by four main themes in the Table II. Almost similar trends with the analysis of weekly online learning reports was identified in the social network monitoring.

TABLE I. TEMATIC ANALYSIS OF ONLINE LEARNING WEEKLY REPORT THAT STARTED TO COLLECT FROM EVERY TEACHER AFTER OPDP DELIVERY.

Teacher	• Lack of equipments (headphone, camera etc…) for online learning • Increased workload in online learning in comparison with traditional face-to-face learning Health problems (ear pain, low back et…) related long time working online.
Student	• Students asked to record online lectures that can be reviewed due to internet access problem • Use of mobile devices to the online learning decreased importance of practical courses such as seminar and laboratory and it is being similar to lectures. • Due to high cost and bad internet connection (especially for student from countryside), student engagement and motivation to join courses were decreased dramatically. • High cost of data usage to connect MS Teams program that we adapted as an online learning tool after OPDP
Management	• The time overlap of online courses and other meetings such as department meeting, management team meetings, human resource training, and thesis defense • Complain about solely use of old LMS and neglecting the use of newly adopted MS-team for online learning quality assurance and teachers workload assessment mechanism.

TABLE II. ANALYSIS OF TEACHER'S SOCIAL NETWORK AFTER DELIVERY OF OPDP

1. Posts related class management of the online learning
• Where is my class? Who create course team at newly adapted MS teams program? • Where are my students? Who register students to the course team at MS teams? • Student id number overlapped and have a problem to register to the course team
2. Post related teacher's initiatives and experience sharing
• Video tutorials for course team creation, adding student to the course team • Manuals for scheduling course set-up • Use of digital content, content creation, live streaming tips • Introduction to create MOOC etc.
3. Posts reporting challenges related internet speed, stability and high cost for data usage
• Students from countryside were inactive due to bad internet connection • Student complaining about high cost of data usage for newly adopted MS Teams platform • Student engagement and motivation in online learning decreased due to high cost of data usage
4. Posts related to the time management and quality assurance
• Complain about neglecting the use of newly adopted MS-team for online learning quality assurance and teachers workload estimation • Overworking in online format than traditional face-to-face teaching • Time overlapping issue between courses and non-teaching activities such as management meetings, thesis defense etc.

IV. DISCUSSION

The current study was purposed to share our experience of implementing first ever online professional development program (OPDP) at university level that can leverage paradigm shift of online learning due to COVID-19. More detailed information of planning and design, implementation and delivery of the OPDP was discussed in materials and method section. Thus, discussion section will be more likely focused on the program evaluation and post-training assessment results.

The teacher's satisfaction survey for OPDP was analyzed using SURE model [23, 24] (Fig. 2). It demonstrated that the training program had high quality contents (K_{sure}=0.96), best management and planning (K_{sure}=0.98), and sufficient participant engagement (K_{sure}=0.94). However, participant responded they could not acquire all relative knowledge and skills provided in the ODPD (K_{sure} ranged between 0.81 - 0.84). It seems that two week of training time was not enough to acquire necessary knowledge and skills that we purposed to be given. Moreover, the result indicates teacher's self-motivation to engage in the program (K_{sure}=0.92) was relatively low (Fig. 2). Further professional development programs needs to enhance teacher's motivation. One way to enhance teacher's engagement and motivation may link the compensation mechanism to the professional development program [25]. During OPDP delivery, low internet speed and stability (K_{sure}=0.81) was challenging issue. This is not only drawback of ODPD, but also main problem observed in developing country [26, 27].

One of the important outcome of OPDP is that MUST took campus wide digital transformation by utilizing MS Teams platform for online learning tool. Form viewpoint of teacher, it was best action to accelerate online learning, though such sudden shift brings several challenges to other stakeholders especially for the students (Table I and II). The biggest problem for student was the high cost of the data usage when using MS-Teams platform as an online learning tool. Such high cost also led to decreased course engagement and student motivation to join online courses (Table I). It was our mistake that we did not negotiate with internet enterprises [7] to find way to decrease data cost before deciding to use MS teams as a campus wide online learning tool. The result clearly indicated that it is important to account all stakeholder's benefit and loss before making university level digital transformation.

Our result implies there is big problem related workload arise in online learning (Table I and II). Teachers reported increased workload in online learning than traditional face-to-face teaching. International study demonstrated the workload is at least 20% higher in online learning in comparison of face-to-face format [28]. Thus, the policies related workload management and compensation mechanism in online learning must be revised based on sufficient research data for Mongolian context. Result indicated there is time overlap between online courses and other non-teaching online meetings. This implies the e-learning calendar of the MUST was not complete enough and more likely covers on teaching oriented activities. There is need of comprehensive digital calendar that covers not only online courses but also research and management activities. The calendar can be created and shared using more open and digital format by using Google or MS-Outlook calendar.

Table II revealed online course management at MS-teams platform that we adopted after OPDP, such as creating subject groups or adding respective students, was still not clear to the professors. More systematic guideline should be provided and there is need of special team for handling such task rather than rely on teachers only. The issue of online learning evaluation and quality assurance mechanism remain not suitable (Table II). Since online learning management system was replaced by MS-teams platform, old policy that require to utilize previous learning management system need to be renewed.

Although there were several negative outcomes after implementation of OPDP and campus wide digital transformation, there is a number of positive trends and results. For instance, teacher's self-motivation and contribution to online learning increased and negative attitude toward change was decreased. It can be shown from teacher satisfaction survey (Fig. 2) and social network posts at teacher's Facebook group (Table II). After delivery of OPDP, initiatives to work with OEC related online learning, such as co-development of MOOC and OER were significantly increased. The study has several limitations. The follow-up research of qualitative analysis of the weekly report and social network monitoring only conducted for teachers at main campus located in Ulaanbaatar. Which means there could be different trends, problems and needs to enhance online learning for other campuses located in countryside region. The MOOC developed was used for source material for flipped classroom during OPDP implementation. To increase the impact of the MOOC on sustainable professional development, localized edx platform (www.edx.edu.mn) at MUST needs more professional upgrade.

V. Conclusion

The study shared the best experience of developing and implementing "online professional development program" during COVID-19 pandemic for Mongolian higher education. Mass teacher engagement to the program implied it is not national best example but also international benchmark for organizing online professional development training. The program delivery success was 95%. A number of important observations to enhance quality of online learning, especially for developing country, were derived from post-training follow-up research and discussed through the study.

Acknowledgment

This work was supported by DeSTT – "Development of Skills and Teachers Training for Leadership" project (609905-EPP-1-2019-1-EPPKA2-CBHE-JP) co-funded by the Erasmus+ Programme of the European Union

References

[1] T. Hale, & S. Webster, (2020). "Oxford COVID-19 government response tracker,".

[2] J. Crawford, K. Butler-Henderson, J. Rudolph, B. Malkawi, M. Glowatz, R. Burton, & S. Lam, (2020). "COVID-19: 20 countries' higher education intra-period digital pedagogy responses," Journal of Applied Learning & Teaching, 3(1), 1-20.

[3] UNESCO: ICT in Education Policy Review in Mongolia. (2021).

[4] Challenges and responses to COVID-19 in higher education: Perspectives from Asia-Pacific. (2020).

[5] N. Almazova, E. Krylova, A. Rubtsova, & M. Odinokaya, M. (2020). "Challenges and opportunities for Russian higher education amid COVID-19: Teachers' perspective," Education Sciences, 10(12), 368.

[6] C. Carrillo, & M. A. Flores, (2020). "COVID-19 and teacher education: a literature review of online teaching and learning practices," European Journal of Teacher Education, 43(4), 466-487.

[7] W. Ali, (2020). "Online and remote learning in higher education institutes: A necessity in light of COVID-19 pandemic," Higher Education Studies, 10(3), 16-25.

[8] O. Zawacki-Richter, (2021). "The current state and impact of Covid-19 on digital higher education in Germany," Human Behavior and Emerging Technologies, 3(1), 218-226.

[9] S. Donitsa-Schmidt, & R. Ramot, (2020). "Opportunities and challenges: teacher education in Israel in the Covid-19 pandemic," Journal of Education for Teaching, 46(4), 586-595.

[10] M. J. Sá, & S. Serpa, (2020). "The COVID-19 pandemic as an opportunity to foster the sustainable development of teaching in higher education," Sustainability, 12(20), 8525.

[11] L. Sokal, L. E. Trudel, & J. Babb, (2020). "Canadian teachers' attitudes toward change, efficacy, and burnout during the COVID-19 pandemic," International Journal of Educational Research Open, 1, 100016.

[12] S. Tejedor, L. Cervi, A. Pérez-Escoda, & F. T. Jumbo, (2020). "Digital literacy and higher education during COVID-19 lockdown: Spain, Italy, and Ecuador," Publications, 8(4), 48.

[13] N. Kaur, (2020). "The Face of Education and the Faceless Teacher Post COVID-19 Naginder Kaur1 and Manroshan Singh Bhatt2," Horizon, 2, 39-48.

[14] T. Beteille, E. Ding, E. Molina, A. Pushparatnam, & T. Wilichowski, (2020). "*Three Principles to Support Teacher Effectiveness during COVID-19,"*

[15] J. Whalen, (2020). "Should teachers be trained in emergency remote teaching? Lessons learned from the COVID-19 pandemic," Journal of Technology and Teacher Education, 28(2), 189-199.

[16] R. Hartshorne, E. Baumgartner, R. Kaplan-Rakowski, C. Mouza, & R. E. Ferdig, (2020). "Special issue editorial: Preservice and inservice professional development during the COVID-19 pandemic," Journal of Technology and Teacher Education, 28(2), 137-147.

[17] J. Reynolds-Cuéllar, G. S. Stump, & A. Bagiati, (2020, October). "Building a Community for Educational Transformation in Higher Education," In 2020 IEEE Frontiers in Education Conference (FIE) (pp. 1-5). IEEE.

[18] T. Karalis, & N. Raikou, (2020). "Teaching at the times of COVID-19: Inferences and Implications for Higher Education Pedagogy," International Journal of Academic Research in Business and Social Sciences, 10(5), 479-493.

[19] M. W. Marek, C. S. Chew, & W. C. V. Wu, (2021). "Teacher experiences in converting classes to distance learning in the COVID-19 pandemic," International Journal of Distance Education Technologies (IJDET), 19(1), 40-60.

[20] S. K. Wang, & H. Y. Hsu, (2008). "Use of the webinar tool (Elluminate) to support training: The effects of webinar-learning implementation from student-trainers' perspective," Journal of interactive online learning, 7(3), 175-194.

[21] M. B. Gilboy, S. Heinerichs, & G. Pazzaglia, (2015). "Enhancing student engagement using the flipped classroom," Journal of nutrition education and behavior, 47(1), 109-114.

[22] C. Chen, G. Sonnert, P. M. Sadler, D. D. Sasselov, C. Fredericks, & D. J. Malan, (2020). "Going over the cliff: MOOC dropout behavior at chapter transition," Distance Education, 41(1), 6-25.

[23] U. Tudevdagva, (2014). "Structure oriented evaluation model for e-learning,"

[24] U. Tudevdagva, (2020). "Online Tool of Data Processing for the SURE Model. In Structure-Oriented Evaluation (pp. 77-87)," Springer, Cham.

[25] F. T. Gemeda, & P. Tynjälä, (2015). "Exploring teachers' motivation for teaching and professional development in Ethiopia: voices from the field," Journal of Studies of Education, 5(2).

[26] B. Mundial, (2020). "Remote Learning and COVID-19: The use of educational technologies at scale across an education system as a result of massive school closings in response to the COVID-19 pandemic to enable distance education and online learning," Versão de, 16.

[27] H. A. Z. Rony, & S. T. Awal, (2019). "University Teachers' Training on Online Teaching-Learning using Online Platform during COVID-19: A Case Study," Bangladesh Education Journal, 57.

[28] W. L. Worley, & L. S. Tesdell, (2009). "Instructor time and effort in online and face-to-face teaching: Lessons learned," IEEE Transactions on Professional Communication, 52(2), 138-151.

The Evaluation of Online Yoga Course during Pandemic Period

Uranchimeg Tudevdagva
Faculty of Computer Science
Chemnitz University of Technology
Chemnitz, Germany
uranchimeg.tudevdagva@imformatik.tu-chemnizt.de

Zandi Nyamjav
Department of Traditional Therapy
International school of Mongolian Medicine
Ulaanbaatar, Mongolia
zandi@mnums.edu.mn

***Abstract*—This paper describes findings from evaluation of online yoga course during lock down in Mongolia. The course was offered open in public. Any person with any age, educational background or professional was possible to enroll to offered course. Main content of course targeted not only to teach yoga exercises, moreover focused to teach how to care liver and clean the liver at home. By other definition this was combined course of yoga and health care. This specific course offered three times from March to May 2021. In three session of online course enrolled more than 300 learners from not only Mongolia as well as from other countries: Germany, Austria, USA, Philippines, and Korea for instance. After each online course we collected data from leaners via online survey. In total 149 learners took part of the online survey.**

Keywords—health care, liver, self cleaning, SURE model, evaluation goal, goal structures.

I. Introduction

Yoga is one of the most commonly used mind-body interventions [1-3]. Recent research suggests that mind-body practices such as yoga may have beneficial effects on health. COVID-19 has emerged as a chief stressor among the general population [4-6]. Therefore, there is a need for stress management tools, such as Yoga, breathing exercises and meditation, both at home and at work. These can be adopted during the pandemic with proper maintenance of social distancing.

Due to the Covid 19 virus the world is changed dramatically in last one and half year [7-10]. Flights are stopped, schools, fitness centers, swimming pools, cinema theaters, operas, shops, and restaurants are closed. Many countries announced long and strong lock downs. People must was sit in home and to fit to new situations. This was unexpected shock for all nations of all countries.

II. Evaluation process

A. Purposes of evaluation process

To support own nation during lock down period we decided to offer online yoga course with combination of health contents. The program of online yoga and health course developed based on long years experience of yoga trainer/master teacher and doctor of traditional medicine. The founder of this online yoga and health course did many times yoga courses face to face before 2020. She had collected enough experience during 20 years of yoga teaching in different level and reached as master in yoga training courses in India.

Since November of 2020 she offered online yoga and health course to public. This course was first test for such kind of combination course together with yoga and health courses. Courses from November and December 2020 were successful and participants of courses are asked to repeat or continue such us courses to offer publically. Therefore, from January of 2021 designed new courses with duration of 21 days for each course and offered them publically by Facebook channel usually.

Main of aims of this evaluation is to figure out next points from as feedback from participants:

- The understanding of participants relating to lecture which is covering health content
- Improvement of health knowledge of participants by giving lecture
- Quality of yoga exercises and meditation course
- Knowledge of healthy style living
- Visible results after course
- Quality of online teaching, skills of tutor

Above listed six key aims defined by questions. First four aims are essential important for evaluators. Target of combined online course was to help people to over go hard lock down easily and useful. To improve their information and basic knowledge about yoga, medication and healthy style life. Therefore, by this self-evaluation process we expecting to find out how good we reached our defined targets.

B. The SURE model

The structure-oriented evaluation (SURE) model originally was developed for e-learning with focus of multi-dimensional evaluation spaces [11-14]. Core of the SURE model is logical structures. There are two different logical structures can be applied. The series structures are focused to visualize key goals of evaluation process. The parallel structures are focused to visualize sub goals of evaluation process. By other words, here in the SURE model most important is to define evaluation goals and visualize them via logical structures: series and parallel.

III. Implementation of the SURE model

For implementation of the SURE model evaluators have to follow pre denied eight steps.

A. First step.

In this step evaluators have to define key goals of evaluation. In this evaluation case we defined five key goals:

- B_1 – Lecture content and quality

- B_2 – Health knowledge of participants
- B_3 – Adoption of proper habits
- B_4 – Quality of yoga exercises and meditation course
- B_5 – Quality of online teaching, skills of tutor
- B_6 – Training outcome

By the SURE model rule, general evaluation will successful only if all these five key goals are evaluating as successful, or bigger than zero (>0).

B. Second step.

Here evaluators have to define sub goals which in sense of support to reach key goals their goal successful. That means each key goal should have at least one sub goal. In general, there are not hard limit to up number but by previous evaluation cases experience should not be go over ten sub goals for single key goal. Below is list of sub goals:

B_1 – Lecture content and quality

- Subject of lecture fit to content
- Content of lecture was easy to understand and clear
- Duration of lecture was exact
- Support from lecturer helped to understand lecture well
- Question session after lecture was very useful

B_2 – Health knowledge of participants

- My knowledge about healthy food was improved
- I got knowledge about healthy life
- My knowledge about basic background for healthy life is improved
- My basic knowledge about healthy life is improved
- The knowledge which I got from this course distributed among my family members

B_3 – Knowledge of healthy lifestyle Adoption of proper habit

- I am getting up early now then before
- Sleeping early than before
- Taking healthier food
- Prone to regular active exercise
- Having better mood and more positive

B_4 – Quality of yoga exercises and meditation course

- Yoga exercises from trainer performance was excellent
- Guidance from trainer was good
- Exercises was easy to follow
- Result of exercise was noticeable to me
- To end of course my performance improved

B_5 – Quality of online teaching, skills of tutor

- Tutor's preparation for the classes was good
- Teaching method, outlook and content delivery were good
- Tutor's oratory was good
- Tutor's performance in techniques was good
- Tutor was responsible and able to deliver the targeted info

B_6 - Training outcome

- My health condition has improved
- Became more positive and enthusiastic
- Sleep quality has improved
- Could adopt the habits of healthy lifestyle
- Became able to give health tips to family members and friends

Figure 1 shows logical structure of sub goals. There are 5 sub goals for each key goal defined.

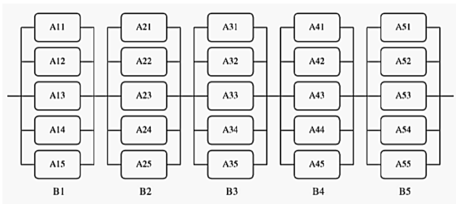


Fig. 1. The logical structure for sub goals

C. Third step.

After complication of first two steps logical structures of evaluation should be defined. The defined logical structures of evaluation have to be checked and proofed by evaluation team, or groups who can be interested or applied evaluation results after evaluation process. Result of this step should be writing protocol about confirmation of logical structures of evaluation.

In our case the evaluation team consist from two individuals: yoga master and SURE expert. And both confirmed defined evaluation goals easily, because the logical structures of evaluation goals were developed in cooperation of these two persons.

D. Fourth step.

In this step have to create adapted checklist (questioner, survey) for data collection based on defined sub goals. By other words based on definition of sub goals have to formulate question for data collection.

E. Fifth step.

In this step created checklist have to be checked and accepted by evaluation team. The Google form created by SURE expert and yoga master checked and confirmed the questions.

F. Sixth step.

In this step by created checklist have to be collected data. In our case we distributed online link to all participants and they are attended to survey voluntarily. Data collection continued in a week that means shared link was open for one week for participants.

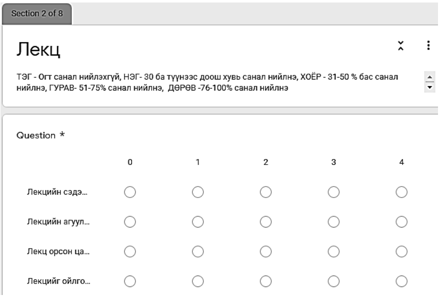


Fig. 2. The example of questions by Google form

We applied Google form for data collection and in Figure 2 shows an example of questions.

G. Feventh step.

Here collected data have to be processed by SURE calculation rules.

H. Eight step.

The evaluation scores of the SURE model delivered as table from the online calculator.

IV. RESULTS OF THE SURE EVALUATION

After the data processing the SURE model delivering four different evaluation scores. Below describe all evaluation scores in separated sub paragraphs.

A. General evaluation score

The online calculation tool computes all collected data and by main formula of the SURE model and as general result of evaluation delivers main evaluation score. This single main score calls general evaluation score or SURE score. In Table 1 shows general evaluation scores of all three evaluations.

TABLE I. SURE EVALUATION SCORES

Months	**General Evaluation score**	**Sample size**	**Gender**
March	0.85	39	Female 39
April	0.90	54	Female 52, Male 2
May	0.88	53	Female 50, Male 3

B. Evaluation scores for key goals

In first step of the SURE model evaluators defined key goals of evaluation process. In Table 2 shows all evaluation scores for key goals in all three evaluation processes.

TABLE II. EVALUATION SCORES OF KEY GOALS

Key goals	**B_1**	**B_2**	**B_3**	**B_4**	**B_5**
March	0.88	0.85	0.82	0.86	0.88
April	0.92	0.91	0.89	0.92	0.91
May	0.89	0.87	0.86	0.89	0.91

C. Evaluation scores for sub goals

In step two evaluators defined sub goals to support acheivement of key goals. In Table 3 shows evaluation scores of sub goals.

TABLE III. EVALUATION SCORES OF SUB GOALS

Sub goals	**March**	**April**	**May**
A_{11}	0.88	0.91	0.89
A_{12}	0.88	0.91	0.89
A_{13}	0.88	0.88	0.89
A_{14}	0.88	0.91	0.89
A_{15}	0.87	0.89	0.88
A_{21}	0.84	0.9	0.86
A_{22}	0.84	0.9	0.87
A_{23}	0.85	0.9	0.85
A_{24}	0.83	0.89	0.86
A_{25}	0.77	0.83	0.82
A_{31}	0.79	0.81	0.83
A_{32}	0.79	0.82	0.79
A_{33}	0.79	0.86	0.83
A_{34}	0.78	0.85	0.85
A_{35}	0.81	0.88	0.86
A_{41}	0.87	0.91	0.89
A_{42}	0.86	0.91	0.88
A_{43}	0.85	0.92	0.89
A_{44}	0.84	0.89	0.88
A_{45}	0.83	0.89	0.88
A_{51}	0.88	0.91	0.91
A_{52}	0.88	0.9	0.91
A_{53}	0.88	0.91	0.91
A_{54}	0.88	0.91	0.91
A_{55}	0.88	0.91	0.92

D. *Evaluation scores of each responces*

The SURE evaluation computes data of each response and calculates evaluation score for every participant. In Figures 1-3 show some part of responses.

	B_1					B_2					B_3					B_4					B_5					
k	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{41}	A_{42}	A_{43}	A_{44}	A_{45}	A_{51}	A_{52}	A_{53}	A_{54}	A_{55}	$Q^*_{c,k}(C)$
1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
3	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	0.89
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
5	4	4	4	4	4	3	3	3	2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	0.87

Fig. 3. First five responses of March

20	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
21	3	3	3	3	2	3	3	3	3	1	2	2	2	2	2	3	2	2	2	3	2	2	2	2	2	0.6
22	4	4	4	4	4	4	4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	1
23	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
24	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
25	4	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1

Fig. 4. Responces from 20 to 25 of April

49	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
50	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
51	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
52	4	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1
53	3	3	3	3	3	2	2	2	2	1	2	1	2	2	3	3	3	3	3	3	4	4	4	4	4	0.67

Fig. 5. Last five responces of May

From figures cannot see numbers clear. But from figures can see color variance clearly. This is one of advantage of the SURE model result table by onlin calculation. Green color means evaluation scores are very high nearly to 1.0. If colors are more yellow means evaluation scores are middle or below than middle. If colors are red it confirms that evaluation scores rea nearly to minimum: 0.0. From all three fogures we can see mainly intensive green colors. That means single participants evaluation was high nearly to 1.0.

V. Statistical analyze of data

Total number of three times (March, April and May) data collection was 146. This is 48.66% of the all participants from three months. There are 5 male participants responded to online survey and it means around 3% of all participants who attended to survey was males. Age difference was from 16 upto 65.

The professional background of participants was very wide: teachers, engineers, architectures, cooks, medical doctors, youga master and trainers, layers The professional background of participants was very wide: teachers, engineers, architectures, cooks, medical doctors, yoga master and trainers, lawyers, nurses, businessman and businesswomen, managers, designers, university lectures, conductor, banker, and translators.

By online learning experiences, in March there are 17 participants attanded first time, 4 participants attended second time, 9 participants attended three times, 3 participants attaned four times and 6 participants attanded more than five times, In April: there are 36 participants attanded first time, 12 participants attended second time, 2 participants attended three times, one participants attaned four times and 3 participants attanded more than five times. In May: there are 38 participants attanded first time, 9 participants attended second time, 3 participants attended three times and 3 participants attanded more than five times.

The technical devices which used for online courses we focused on desktop computers, notebooks, tablets and smart phones. 11 participants used desktop computers, 54 participants used notebook, 15 participants used tablets and 120 participants used smat phones. Total of used devices are more than number or responcses due to the overlapped answers. Some participants used not only device for online courses and thise influenced to amount of used devices.

There are 130 participants used home internet or fixed modem internet connections for online course and 37 participants used mobile data from smart phones.

VI. Conclusion

Total number of three times (March, April and May) data collection was 146. This is about 48.66% of all participants during three months. That means we cannot say 100% probablity to reaction of whole participants by collected data. However, by 146 responces in all three months general evaluation score of online course was between 0.85 up to 0.90. It shows that success of online course was always bigger than 80% from 100%. The SURE evaluations scores measure sclae is always between 0 and 1. It easy to understand and explain. Tables 1-3 contains all evaluation scores which calculated by online calculation tool of the SURE model.

There are no any zero (0) value and this confirms that all defined goals are reached its target successfully. And all evaluation scores are bigger than 0.70, by means acheivement of online course fullfilled with quiate high evaluation scores.

This self evaluation made first time by yoga master for her online courses. Main purpose of this evaluation was to figure out success of defined goals by yoga master due to the combination of yoga exersices together with traditional medicines information with target to imptove ans upport public knowledge of health and helathy style life.

By evaluation scores we can conclude that public online yoga and health courses reached their target succussfully. At least half of the total participants satisfied by offered content and performance of course.

References

[1] A. Hiranmayananda, Cure yourself with yoga. New Delhi, 2005

[2] S.S.Ravi Shankar, Commentary on Patanjali yoga sutras. Sri Sri publications trust, India, 2010

[3] S.W.Devananda, The Sivananda companion to yoga. Fireside rockfeller center, NY, 2010

[4] Nagarathna R, Anand A, Rain M, Srivastava V, Sivapuram MS, Kulkarni R, Ilavarasu J, Sharma MNK, Singh A and Nagendra HR (2021) Yoga Practice Is Beneficial for Maintaining Healthy Lifestyle and Endurance Under Restrictions and Stress Imposed by Lockdown During COVID-19 Pandemic. Front. Psychiatry 12:613762. doi: 10.3389/fpsyt.2021.613762

[5] The Lancet. India under COVID-19 lockdown. Lancet. (2020) 395:1315. doi: 10.1016/S0140-6736(20)30938-7

[6] Pietrobelli A, Pecoraro L, Ferruzzi A, Heo M, Faith M, Zoller T, et al. Effects of COVID-19 lockdown on lifestyle behaviors in children with obesity living in Verona, Italy: a longitudinal study. Obesity. (2020) 28:1382–5. doi: 10.1002/oby.22861

[7] Kimberly Chriscaden, Impact of COVID-19 on people's livelihoods, their health and our food systems, Online available:

https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems

[8] Chaturvedi, K., Vishwakarma, D. K., & Singh, N. (2021). COVID-19 and its impact on education, social life and mental health of students: A survey. Children and youth services review, 121, 105866. https://doi.org/10.1016/j.childyouth.2020.105866

[9] Aucejo E.M., French J., Ugalde Araya M.P., Zafar B. The impact of COVID-19 on student experiences and expectations: Evidence from a survey. Journal of Public Economics. 2020;191 doi: 10.1016/j.jpubeco.2020.104271

[10] Hasan N., Bao Y. Impact of “e-Learning crack-up” perception on psychological distress among college students during COVID-19 pandemic: A mediating role of “fear of academic year loss” Children and Youth Services Review. 2020;118 doi: 10.1016/j.childyouth.2020.105355.

[11] Uranchimeg Tudevdagva, and W. Hardt, A measure theoretical evaluation model for e-learning programs. In Proceedings of the IADIS on e-Society 2012, March 10-13, 2012, Berlin, Germany, pp.44-52.

[12] Uranchimeg Tudevdagva: Structure Oriented Evaluation Model for E-Learning. Wissenschaftliche Schriftenreihe Eingebettete Selbstorganisierende Systeme, Universitätsverlag Chemnitz, Chemnitz, Germany, July 2014. 123 p., ISBN: 978-3-944640-20-4, ISSN: 2196-3932.

[13] Uranchimeg Tudevdagva, W. Hardt and L. Bayar-Erdene, The SURE Model for Evaluation of Complex Processes and Tool for Implementation. The 9th International Forum on Strategic Technology (IFOST 2014), IEEE Computer Society, October 21-23, 2014, Chittagong University of Engineering and Technology, Bangladesh.

[14] Uranchimeg Tudevdagva, “Structure-Oriented Evaluation An Evaluation Approach for ¬Complex Processes and Systems”, Gewerbestrasse 11, 6330 Cham, Switzerland, Springer, 2020, pp.92, ISBN 978-3-030-44805-9 ISBN 978-3-030-44806-6 (eBook), https://doi.org/10.1007/978-3-030-44806-6.

The Evaluation of Adult Training Courses During the COVID-19 Quarantine

Uranchimeg Tudevdagva
Faculty of Computer Science
Chemnitz University of Technology
Chemnitz, Germany
utu@hrz.tu-chemnizt.de

Enkhtsetseg Luvsandagva
Centre of lifelong education
Darkhan-Uul, Mongolia
enkee1213@gmail.com

Altantsetseg Surenjav
Centre of lifelong education
Darkhan-Uul, Mongolia
s.altantsetseg11110@gmail.com

***Abstract*—This paper describes evaluation results of six different adult training courses during COVID-19 quarantine in Darkhan-Uul aimag, Mongolia. Trainings are focused on adults and offered from lifelong learners' organization. Aim of the evaluation process is to compare online trainings with traditional classroom training, to find strengths and weaknesses of the online training to improve online training in future, to develop efficiency of offered courses in next rounds. Further aims are, the evaluation process should provide theoretical knowledge about evaluation process to course teachers. Teachers of lifelong learner center should collect experience by this evaluation process. Applied methods was online survey and collected data processed by standard statistical processes. In total 122 adults took part in online survey.**

Keywords—pandemic, adult education, continues education, lifelong learner.

I. INTRODUCTION

The concept of lifelong learning was first introduced by UNESCO in 1972 in Edgar Faure's Learning to Be Human: Education Today and Tomorrow [1]. Jacques Delor's 1996 book, Education: The Hidden Treasure [2], addresses the four pillars of lifelong learning and learning. The Organization for Economic Co-operation and Development (OECD) published a report entitled "Lifelong Learning for All" in 1996, warning that education should be considered together to address the economic, social, cultural and environmental challenges facing society. Since 2000, the European Union has held a wide-ranging discussion on lifelong learning as part of its goal to create a knowledge economy [3].

In 2006, the European Parliament supported the Lifelong Education curriculum for 2007-2013 [4].

India's National Skills Initiative focuses on the development of specific skills for lifelong learning, while Malaysia's 2011-2020 Plan for education of Lifelong focuses on adult education and training [5].

In 1995, the Law on Education stated that "Mongolia's education system is a combination of formal and non-formal education," assuming that the Mongolian government will meet the lifelong learning needs of its citizens by improving the combination of formal and non-formal education.

In 1997, the Government of Mongolia approved the "National Program for the Development of Non-Formal Education". The Law on Education of 2002 states that "Citizens may receive education through formal and informal training".

In 2010, the Government of Mongolia and the Ministry of Education, Culture and Science approved the "Policy on Non-Formal Education" [6].

Darkhan-Uul aimag's Citizens' Representative Khural's Resolution No. 78 of 2011 and aimag Governor's Decree No. 591 of 2011 established the NFE as a "Civic Education Center" and started its operation on January 2, 2012 [7]. Aimag Governor's Decree No. 407 of November 20, 2013 extended it into a "Lifelong Education Center" to provide education to families, citizens, moral development, aesthetics, life skills and all levels of society needed to improve the quality of life. Activities are aimed at providing literacy, basic and complete secondary education to children and young people who have no out-of-school education [8]. Below is training directions for lifelong education for learners:

- Life skills
- Family
- Moral-Maturity
- Civil education
- Aesthetic-sensory education
- Education for sustainable development
- Financial education

The Education Compensation Program provides training, advocacy, and advocacy activities based on the needs of citizens, organizations, and specific target groups (the elderly, children, youth, people with disabilities, etc.) in the areas of basic and upper secondary education.

The training forms of the Centre of lifelong education:

- Classroom or equivalent
- Distance
- Mobile or home training

Training and advocacy activities were conducted in 5 areas of lifelong education, which increased public awareness of lifelong education. A total of 3000 people participated in the training [9].

II. EVALUATION METHODOLOGY

The online survey is selected as the main methodology for evaluation process. Survey questions are formulated as result of analyze of course content and aim, discussion with evaluation expert and teachers for training courses.

Below are key focuses of evaluation questions:

- Course contents;
- Practical trainings;
- Skills of teacher;
- Quality of online lessons.

Above listed focuses are becoming basics for questions with selection options. Questions are formulated as statements and participant should select one of offered options as answer to question. There are five different answer options which shows how much agree with statement:

- Not agree at all;
- Agree up to 30%;
- Agree between 31% - 50%;
- Agree between 51% - 75%;
- Agree up to 100%.

Beside defined closed questions some of open questions are included into survey:

1. What are the benefits and importance of the training course?
2. What are the advantages of traditional classroom training?
3. What are the disadvantages of traditional classroom training?
4. Your opinion about training

III. Statistic Data

Totally attended 122 participants to six different trainings.

- There are 88 females (72.1%) and 34 males (27.9%), see Fig.1.
- Age difference was between 18 and over 55. Most part took younger participants between 15 and 35, see Fig.2.

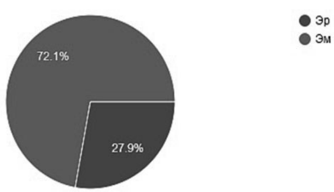


Fig. 1. Genders of participants (эр – males; эм - females)

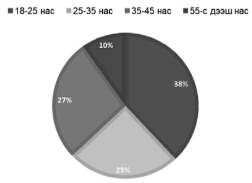


Fig. 2. Data for ages (нас – age, дээш нас – age up)

IV. Results of Evaluation

The first focus of training course evaluation consists of five questions and targeted to figure out feedback on content and teaching methods:

- The content of the training was clear
- Trainings' target was clear
- Teacher of training used scheduled time fully for training
- Some active learning methods are applied for training
- My theoretical knowledge is increased by training

As evaluation result 65 participants selected highest score (Agree up to 100%) from offered answers (see Fig.3). Four participants don't agree at all on giving statements.

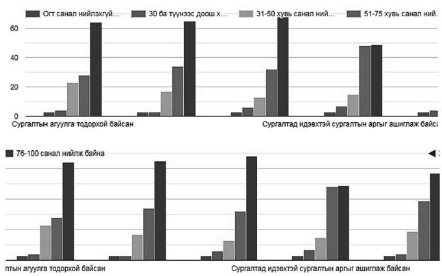


Fig. 3. Evaluation results for training course

- Improved my theoretical knowledge by practical training
- My skills are added and improved by practical training
- Practical trainings run learner-based methodology

- Practical training targeted to teach learners to understand giving problems and find corresponding solutions

The question “Improved my theoretical knowledge by practical training” evaluated by 47 learners as agree between 51-75%. The question “My skills are added and improved by practical training” evaluated by 52 learners as agree between 76-100%. The question “Practical trainings run learner-based methodology” evaluated by 49 learners as agree between 76-100%. The question “Practical training targeted to teach learners to understand giving problems and find corresponding solutions” evaluated by 54 learners as agree between 76-100%.

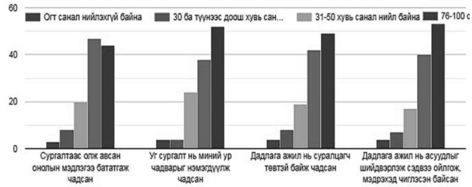


Fig. 4. Feedback on practical trainings

- Teacher well prepared to training
- Quality of teaching methods was excellent
- Presenting skill of teacher was excellent
- Skill to use technical tools was excellent

In total 76 participants responded with 76-100% to first question: Teacher well prepared to training. The question “Quality of teaching methods was excellent” evaluated by 64 learners as agree between 76-100%. The question “Presenting skill of teacher was excellent” evaluated by 77 learners as agree between 76-100%. The question “Skill to use technical tools was excellent” evaluated by 76 learners as agree between 76-100%.

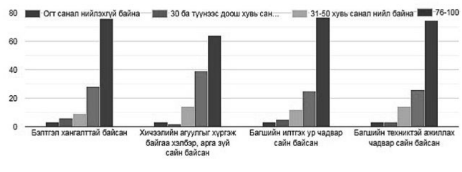


Fig. 5. Teacher skills and quality

- Training with content about family economy and finance was very helpful, I learned a lot of new knowledge about this topic
- Family planning, communication in the family, understand and respect one to others in family such us knowledge and information were very useful
- I attended 6 different trainings in the lifelong learning center and always very satisfied with courses
- Received information about everyday life skills that need to be confident and wealthy, how to take care of yourself also to stop woman violence.
- Got information of sustainable development was included green jobs and sort waste out.
- Make good choice for future jobs, give advice for teenage girls how to behave themselves should care themselves
- Advantage of being determined for your future self development.

V. Conclusion

This evaluation was a first self evaluation process in training courses for life long learners. The main aim of this evaluation is to give knowledge and theoretical understanding about evaluation process to teachers of lifelong learner centers. Morover, to collect experiences on preparation and application of online survey, to learn understand and explain processed data, to write and deliver crossponding reports based on processed evaluation data.

Due to the very first case of self evaluation authors applied no any special evaluation models and data processing. Data results which are included in this work produced by Google form application. The tables and graphics which provided by Google form is easy to understand for beginners in evaluation process. Important in evaluation process is to define clear goal of evaluation. And this essential important issue takes time and discussion between evaluation expert and trainers. Via such dicsions the trainers learned how to define evaluation goals.

In the conclusion this evaluation case helped to collect experience on educational evaluation process for lifelong learner trainers. Self evaluation is useful for trainers to understand not only evaluation process, moreover, opened new view to offering courses from lifelonf learner center to learners. The evaluation process shows some nuances of training and improved the feeling of trainees from prespective of learners. After this first try trainers are willing to repeat the evaluation in coming courses.

References

[1] Faure Report, 1972, Collection of Lifelong Learning Policies and Strategies. UNESCO UIL. [online] Available at: http://uil.unesco.org/lifelonglearning/lifelong-learning-policies-database

[2] Delors, J. et al. 1996. Learning: the treasure within; report to UNESCO of the International Commission on Education for the 21st Century. Paris, UNESCO. European Council. 2002. Council resolution on lifelong learning. [online] Available at: http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2002:163:0001:0003:EN:PDF

[3] European Commission. 2000. Memorandum on lifelong learning. [online] Available at: http://arhiv.acs.si/dokumenti/Memorandum_on_Lifelong_Learning.pdf

[4] European Parliament. 2006. Commission proposal for the introduction of a new action program to promote lifelong learning in the 2007-2013 period. [online] Available at: https://eur-lex.europa.eu/legalcontent/

[5] National program for the development of non formal education

[6] Law on education at: https://www.legalinfo.mn/

[7] Darkhan-Uul aimag's Citizens' Representative meeting Resolution No. 78 of 2011 and Aimag Governor's Decree No. 591 of 2011

[8] Official web site of Lifelong learner center in Mongolia, [online] Available at: https://ncle.edu.mn/

[9] Education page: https://bit.ly/3EXttDC

Conversion of Traditional Course to Online Course-Empirical Evidence on Students Learning Experience

Ummay Ubaida Shegupta
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
ummay-ubaida.shegupta@informatik.tu-chemnitz.de

René Schmidt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
rene.schmidt@informatik.tu-chemnitz.de

Wolfram Hardt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
hardt@cs.tu-chemnitz.de

Abstract— **The necessity of E-learning and E-assessment have become obligatory in the formal institutionalized higher education especially due to the global pandemic. In order to retain the continuity and quality of education, E-learning has turned out to be the single-handed solution for the teaching learning process in the tertiary education system. This absolute need has driven the former traditional seminars and lectures to transfer into online seminar and online lecture. This study addresses the implemented E-learning context by considering the theory of online learning which states about three forms of interactions, student-teacher; student-student and student-content whereas one form of interactions individually or combined with others can bring efficacy in learning process. From this point of view, the learning engagement and students' self-efficacy have been studied in this research to practice full paper. For this purpose, semester wide online learning experience of the postgraduate students from computer engineering discipline have been analyzed along with their E-assessment experience. A course named "Hardware Software Co-design II (HSCII)" which has been taught fully online for the first time ever in summer semester 2020 at "Chemnitz University of Technology", is used as the testbed for this study. At the end of the semester, E-assessment has been conducted in this course to assess students' acquired knowledge. This is one of the unique first-hand experiences of E-assessment both for the faculty and for approximately 61% of the registered students of the course. The central learning management system of the university has been used for providing all the necessary course materials and formative assessment opportunities to the students. Besides, the lecturer has used the central webinar system of the university to conduct synchronous online lecture. Finally, the state of the art online assessment tool of the university has been used to assess students' knowledge gain from this course. Among 122 students, 69 have appeared in the final E-examination and data about their experience have been collected through two surveys. One is about E-learning experience and another is about E-assessment experience. The findings of this study indicate the perception of the learners about their learning engagement and self-efficacy in online learning along with showing their deficit of technical orientation of the E-assessment system which are needed to be considered in this domain specific E-learning and E-assessment scenarios to maximize the learning in future.**

Keywords— Educational Technology [syn: E-learning], Higher education [syn: College, University], Computer-based instruction [syn: Internet-based instruction], Computer engineering, Streaming video, Test format [syn: Exam format]

I. INTRODUCTION

E-learning has got a new dimension of importance since the rise of Covid-19 in 2020 which was not long realized for the past 61 years [1]. United Nation has reimagined E-learning and E-assessment as a resilient solution for the 94 percent of the affected learning spaces worldwide [2]. Among these, higher education system is included which is on the verge of great damage resulting in higher dropout rates in the universities [1].

In order to prevent this in all the 16 states of Germany about 424 higher educational institutions are offering hundreds of online and blended learning programs [3]. While many teachers have heavily invested in online teaching with digital tools, some teachers have demanded the immediate stop of the promotion of digital tools [3]. For example, a group of scholars from the well-respected Ludwig Maximilian University in Munich, with the support of many other professors from other universities, demanded the cancelation of studies for the summer semester (#nichtsemester 2020) [3]. Amidst this situation, "Chemnitz University of technology" has introduced online learning environment for the first time ever in the Computer Engineering department by linking pedagogy, technology, and learner needs.

Terms commonly used for E-learning include online learning, internet learning, distributed learning, networked learning, tele-learning, virtual learning, computer-assisted learning, web-based learning, and distance learning. E-learning or online learning can be defined as the use of the internet to access learning materials; to interact with the content, instructor, and other learners; and to obtain support during the learning process, in order to acquire knowledge and to grow from the learning experience [4]. It allows learners to collapse time and space. However, the instructional design must be planned properly to engage the learner and promote learning [4]. Assessment is essential in any teaching learning process and e-learning is not an

The authors gratefully acknowledge funding by the BMBF (16DHB2106)

exception. For this purpose, e-assessment has been introduced in the online learning environment to measure learner's achievement and progress [5].

This exploration study observes e-learning and e-assessment in a novel setting of the university's teaching learning process and deduces a concept for e-examination. The following section will briefly describe the underpinned theories and research objective. Afterwards, the online lecture setting and e-assessment planning along with study methodology is explained followed by the pilot testing results and final online examination results. Finally, the empirical evidences regarding learning engagement and self efficacy are discussed to establish an initial approach to the implementation of online examinations.

II. Theoratical Framework

E-learning or online learning entails that the learner is at a distance from the tutor or instructor connected by using some form of technology (usually a computer with internet connection) to access the learning materials, to interact with the tutor or instructor and with other learners and therefore some form of support is provided to learners [7]. Likewise online lecture refers as an innovative approach for delivering instruction to a remote learner, using the web conferencing tool as the medium [4]. However, it involves more than just the presentation and delivery of materials using the e-learning tool because the learner and the learning process should be the focus of online learning [8]. Hence it is achieved by interaction for creating a sense of presence and a sense of community for online learners, and for promoting transformational learning [9]. Online learners receive learning materials via e-learning technology referred to as a learning management system. Furthermore, learning tasks are processed through an e-assessment system, which are subsequently personalized and contextualized [4]. In this whole learning process, learners interact with the content, with other learners, and with instructors to test and confirm ideas and to apply what they learn [9]. According to Benson & Samawickrema (2009), the design of the educational experience includes the transactional nature of the relationship between instructor, learners, and content that is important for the learning experience [10]. Additionally, Anderson (2008) explained in his theory of online learning that sufficient levels of deep and meaningful learning can be developed, as long as one of the three forms of interaction (student-teacher; student-student; student-content) is at very high levels [4]. Therefore, the other two forms of interactions could be provided at minimal levels or even eliminated without degrading the online learning experience [4].

The basis of this study stands on the interaction between student and teacher. Consequently, technology mediated interaction for online learning environment has been explored with the concept of learner engagement in the online lecture. Learner engagement is understood through the manifestations of an individual's internal states including both the affect and cognition [11]. It is basically an individual psycho-social process that evolves over time and varies in intensity [12]. The multi-dimensional nature of student engagement includes behavioral, cognitive and emotional components, providing a rich understanding of students' learning experience [12]. As in the online learning environment the student has to interact with digital system, the user engagement scale developed by Heather L. O 'Brien, Paul Cairns and Mark Hall (2018) explains the studied learning scenario [13]. It addresses the concepts of focused attention (FA) that means feeling absorbed in the interaction and losing track of time, the concept of Perceived usability (PU) that refers to negative affect experienced as a result of the interaction and the degree of control and effort expended and the concept of rewarding factor (RW) accumulates endurability, inciting curiosity and feeling involved in the learning experience [13]. It is believed that the ability to engage and sustain engagement in online learning environments can result in positive learning outcomes for e-learning [12].

Students' perception of their online learning experience involves perceived ease of use of the educational technology along with their self-efficacy of the content and behavioral intention in continuation of the learning process [14]. Both the concepts of perceived ease of use and behavioral intention are included as the subscales of "Technology Acceptance Model" [14] which will explain the students' opinion about their attitude towards using e-learning tools. Students' perception about their knowledge gain in the online learning environment is reflected by the thought of their self-efficacy [15].

On the basis of the above-mentioned theoretical framework, the objective of this study is to observe the online learning scenario and figuring out the best possible solution for e-assessment in this setting.

III. Method

A. Study Background

In this study, the data of E-learning and E-assessment experience is analyzed from a course named "Hardware Software Codesign (II)" (HSCII) offered at the department of computer engineering in "X" university mainly to the postgraduate students. During the summer semester 2020, this course has been offered as online course for the first time. The online live lectures occurred via the web conferencing tool "Big Blue Button" and all the lecture materials have been provided in the central learning management system of the university known as "OPAL". Though the students of the university are well-acquainted with OPAL as it is used in every other courses for providing the learning materials and course information, the experience of online lecture for this particular course has been unique both for the lecturer and the students. The provision of chat functions, shared notes, interactive board of the web conferencing tool have been embedded in the online lecture pedagogy. The lecturer has also provided the recorded live lecture to the students to support their learning need. At the end of the semester, the semester final examination has also been organized as online examination. In the following subsection, the detail about the E-assessment will be briefly explained.

B. E-Assessment planning and execution

The occurrence of online examination for HSCII course has been new as well like the online lecture. The lecturer and the course organization team along with the researcher's team have worked for the planning and implementation of the online examination. Since 2012 only traditional format of assessment has been used for this course. So, the planning of implementation of e-assessment included two major planning. One is creation of online examination with auto correction functions and another is to organizing the online test scenario. The university provides the online assessment platform named “ONYX” which is integrated in OPAL. This platform provides the option of automatic correction and fulfills all necessary requirements to create an e-examination in the domain of computer engineering [17]. This is why, ONYX has been used to provide the online examination. For the online invigilation, the web conferencing tool “Big Blue Button” has been used. There were seven invigilators in separate web rooms where the pre-registered students have to join to attend the online examination. Each room have had not more than 15 students. After identifying all the students, the invigilator started the exam in the online exam platform for all. To avoid any misconduct the webcams of the students were kept on for whole examination time. As this has been a first-hand experience for all, a pilot online testing has been conducted before the semester final examination.

C. *E-Assessment format*

The online examination question of the semester final examination includes 22 items having the item difficulty index ranging from p = 0.22 to p = 0.95. All the items have been created in contrast to the difficulty level of traditional examination. Following Bloom’s Taxonomy [16], the question pattern includes knowledge level question at the beginning and gradually presented with application and evaluation related problems. Content-wise theory, calculation, algorithm and programming tasks have been given to assess students’ knowledge gain from the online lecture.

D. *Instrument*

There are two surveys which have been used as instruments to gather student’s opinion. The first survey is used to gather information about the online learning experience. This survey has in total 13 questions including demographic information, 11 close ended questions and one open ended question to acquire feedback about online lecture experience. The 11 questions include questions about students previous online learning experience, technical situation of attending online lectures and five sets of Likert scale questions and two sets of Likert type question about the attribute of the lecturer and online teaching feature. The five Likert scales are- Learner engagement scale with three subscales named “Rewarding Factor” (RW), “Focused Attention”(FA) and “Perceived Usability”(PU); Perceived Ease of Use (PEU) ; Academic Self-efficacy (ASE); Behavioral Intention (BI) and Perceived Enjoyment (PE). The Second survey is used to measure the online assessment experience of the students. This survey contains 14 questions including three open ended question asking students about their difficulties during the online examination, difficulties while responding in the online test platform and seeking their recommendation for the final examination. Rest of the questions are all close ended questions which have inquired students’ technical experience about the E-assessment system and procedures.

E. *Sample*

Initially, there were 122 students who registered to the course and finally 63 students have appeared in the final online examination. Both the surveys are given to all the registered students. Among them only 21 students responded in the survey related to online learning experience and 23 students have responded in the survey related to online assessment experience. The sample includes 68% male students and 32% female students who are mostly international students coming from nine different countries (India,Bangladesh,Iran,Germany,Egypt,Nigeria,Poland,Ghana,and Syria) . Another important characteristic of the sample of this study is that 61% of the students have never attended any online examination before.

F. *Data Collection and analysis*

Using the two surveys quantitative and qualitative data have been collected in this study. The first survey had been conducted during the last week of the online lecture in the summer semester 2020. The survey about the online assessment experience have been conducted right after the piloting of the online examination. From the final e-assessment, the number of the students who have passed and failed have been collected. Additionally, an unstructured interview has been conducted with one of the invigilators of the semester final online examination.

As sample size is small and most of the quantitative data were collected from Likert scale and Likert type questions, descriptive statistics (Mode and Frequency distribution) have been conducted. For the open-ended responses, inductive coding has been used for analysis and for the interview a combination of inductive and deductive coding completed the narrative analysis.

IV. Pilot Test Results

A. *Learner Engagement in the Online lecture*

The analysis of the survey about online lecture experience regarding learner engagement represents the level of agreement of the three subscales- (Shown in Fig.1) Rewarding factor (RWi-iv), Focused Attention (FAi-iii) and Perceived Usability (PUi-ii). Here PU(i) and PU(ii) are reverse coded items. The frequency distribution shows the responses towards agreement and strongly agreement for RW(i), RW(ii), RW(iii), RW(iv), FA(i), FA(ii), FA(iii) and for the reverse coded PU(i) and PU(ii) towards disagreement and strongly disagreement.

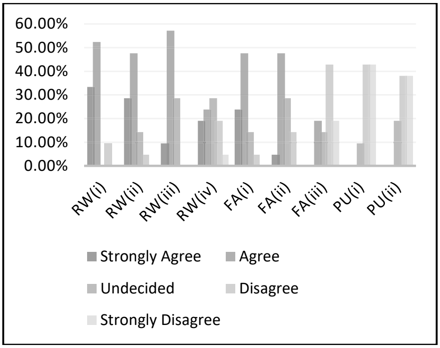


Fig.1. Frequency distribution of learner engagement scale

B. *Perception of learning experience*

Students' perception of their online learning experience have been realized with three scales-Perceived Ease of Use (PEU), Academic Self-Efficacy (ASE) and Behavioral Intention (BI). The mode of the three scales have been presented in the following table.

Table I. Students' perception of the online lecture experience

Scale-wise attributes	Mode (in percentage)	Level of Agreement
PEU(i)	47.62%	Strongly Agree
PEU(ii)	61.90%	Agree
PEU(iii)	57.14%	Agree
PEU(iv)	47.62%	Agree
ASE(i)	66.67%	Agree
ASE(ii)	66.67%	Agree
ASE(iii)	38.10%	Agree
BI(i)	47.62%	Strongly Agree
BI(ii)	47.62%	Agree

The mode of the mentioned attribute in table I shows that most of the students have opined to agreeing all the statements.

C. *Online lecture feature*

Among the six features provided in the online lecture, about 81% students have reported that recording of the lecture is very important to them. Chat functionality is very important to 62% students and to 52.38% students shared notes functionality is very important. Other than that interactive board functionality and video of the lecturer are reported important respectively by 48% and 43% of the students (Shown in Fig.2).

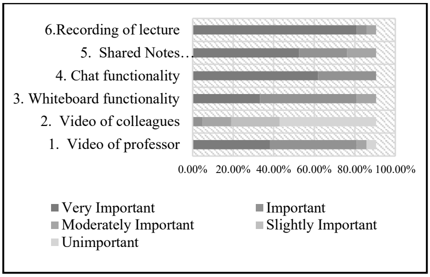


Fig.2.Importance of online lecture features

D. *Perception about lecturer's support*

In the Likert type question responses about lecturer's characteristics and support during online lecture, 47.62% students have responded that the lecturer encourages student participation in online lecture. The response "Strongly Agree" have been got from 47.62% students about the lecturer that the lecturer helps the students to reduce their stress of learning by verbal and body languages in the online lecture. Moreover, 57.14% of the students have strongly agreed that the lecturer used concrete example and visual support to clarify content and 42.86% students strongly agree and another 42.86% students agree that the lecturer is well prepared to take the online lecture. Additionally, 47.62% strongly agree that the lecturer is engaging, responsive and constructive in both tone and content of the speech. A mixed response have been got about using the available features of the web conferencing tool. The "strongly agree" response have been responded by 52.38% whereas 4.76% students disagree and another 4.76% have reported "undecided" about this question.

E. *E-assessment experience*

After the online pilot examination, 41% of the student reported their experience as fair and only 9% reported as excellent (Shown in Fig.3) and to another 14% the online examination experience is good. This pilot testing experience of online examination is reported as poor and very poor in total by 36% of the students.

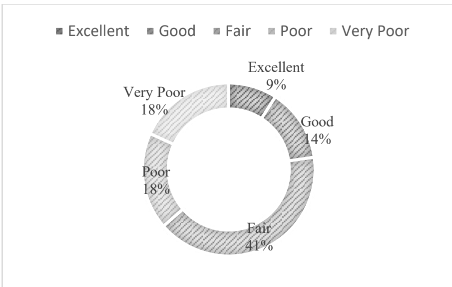


Fig.3.Overall experience of online examination at TUC exam platform

F. *Problems for attending online examination*

Among the seven options provided as problems, about 32% students have responded that they faced technical interruption to be in the web conferencing room during the examination. Another 27.27% students have faced disruption in E-exam platform. About 45.45% students reported that they faced other problems during the online examination which is described in the following qualitative result.

G. *Qualitative result*

For the online lecture of HSCII 67% students have given no feedback while 14.28% students have stated that the online lecture is helping in their learning and want to continue this learning as one of them written, "*I look forward to continue this learning process post-COVID 19. It is novel and a promising experience for possibly better performance. Thank you.*" However, in the online examination about 30% of the student stated problem about the execution pattern of the online examination which includes webcam setting along with attending the examination in the test platform. About 20% of the students have wanted to see a timer to see how much time is left for the examination. Other 8.80% students have reported that they faced technical problems like operating web conference tool and online examination platform in the same device and automatically logging out from the ongoing online examination.

V. Consequences

The previous results show a deficit in the students' use of e-learning technologies due to their limited experience with these technologies since 61% have never attended a online examination before. Likewise, the overall experience was diminished leading to the pilot test experience rating of poor or very poor by 36% of the students.

For this reason, various self-tests have been created in preparation for the final exam. These serve on the one hand to prepare the students for the exams in terms of subject matter and content, but at the same time prepare the students for the different types of examination tasks. Special emphasis have been placed on the diversity of the task types, while at the same time a high taxonomy level of at least level 3 has ensured the needed difficulty level.

Furthermore, the organizational problems like execution pattern encountered in the pilot test were addressed by a clear guide. This guide includes a time line that shows the students the timing of the necessary actions to ensure a smooth process. For this purpose, three phases have been defined. The first phase defines the preparation one or more days before the exam and includes items such as registration on the testing platform, webcam testing, a checklist of necessary materials and equipment. To support the students, a test instance of the conference system has been provided to test all technologies and process steps on their own. In the second phase, the process 30 minutes before the exam start has been described. Here, a step-by-step instruction describes the login on the examination system, the process for authentication and the securing of the online exam situation (e.g. a camera pan through the room). The third phase describes the exam situation itself. Here, all necessary contact information of the respective supervisors have been provided, so that in case of technical problems, help can be requested via other media channels such as e-mail and telephone.

This guide has been made available to the students in parallel on several channels. First, the information have been presented on a web page of the examination system. On the other hand, the information have been sent to all registered students via e-mail.

To address the technical problems, all identified problems were solved or workarounds were found in cooperation with the examination platform provider.

VI. Exam Results

The examination itself has been conducted analogously to the pilot test and used the same types of tasks as the self-tests provided for preparation. All 69 final registered students participated in the exam, so there were no students who could not connect to the exam system or did not know the necessary steps. Subsequently, the results have been evaluated by interviewing the examination supervisors and by analyzing the results of the examination. Criteria of the interview were the analysis of technical problems and organizational problems. A total of 6 examination supervisors participated. None of the interviewees had reported organizational problems during the execution. In total, 3 students with technical problems have been identified, but their problems could be solved directly by the examination supervisors during the examination by e-mail and telephone usage. The data evaluation of the gradings archived shows, about 9.53% students have failed in the exam and 90.47% of the students have passed the online examination. Furthermore, an average grading of 71.97% has been archived.

VII. Discussion

The first-hand experience of institutionalized online lecture and online examination scenario is illustrated as empirical evidence in the present study. In this study, the focus surrounds on the technical elaboration of the execution of the online lecture and the E-assessment. The result about the learning experience of the students have mainly been suppressed by the usage of the educational technology. The qualitative response of the students explained the desirability of the online lecture during this global pandemic time. The first survey results point out the quality of the online lecture corresponding students' perception about engagement in the online lecture, the perceived usability of the tools used in the lecture along with their self-efficacy and intention to attend and revise the video lecture. The findings represent that most of the students had full attention during the online lecture and the content of the lecture has been raising their curiosity. Lecturer's input about participation in the online lecture via the synchronous chat option has made the students felt more involved to this virtual learning environment, as they can ask questions or can clarify confusion about the lecture content. This shows the integration of human teachers' effort with technological equipment that has impacted the e-learning context. Moreover, the findings reflect that the recorded lectures have given the students provision of getting virtual tutoring for revision afterwards and so they have found the recordings as most important feature in the online teaching learning process. Students' responses about lecturer's support show that the lecturer has been putting effort to adapt online teaching pedagogy to facilitate the students learning. This attitude has motivated the students to be resilient for acquiring the domain specific knowledge. Although it is very difficult to retain the concentration in online lectures, the lecturer's continues effort encouraged the students to keep on learning. These findings show in spite of implementation of education technology, the adapted pedagogy of the lecturer has gained students attention, therefore facilitated in learning.

Contrastingly, the survey results about the piloting of e-assessment visualize that students have inquired more about technical enactment of e-examination. Moreover, 36% of the students have reported having online examination experience in the test online examination as unsatisfactory hence identified in the open-ended questions. Nevertheless, 41% students responded the piloting as a "fair" online examination experience, the issues raised have been taken into account for the final online assessment. The results of the interviews of the examination supervisors show the need for good guidance through the process of online examination. The guide as well as communication via redundant media has proven to be an effective method to guide students through the complex process. Likewise, the redundant media channels (phone and mail) have proven to be appropriate methods for agile help during the exam, as all technical problems could be solved directly during the exam. The percentage of the passed students in the online examination is 90.47% which signifies the effort made in the online lecture along with final execution of the online semester final examination.

Furthermore, the high pass rate shows that the lack of experience of the students can be compensated or trained by self-tests which contain tasks assessing higher order learning of HSCII. Additionally, the self-tests cover all necessary test types giving the students a provision of practicing by their own. As a bonus, any misconceptions that arose due to the abrupt switch to the online scenario are compensated as evidenced by the high average score. Altogether, it is visible that incorporating the planning and implementation of e-learning context with e-assessment scenario is important for the successful learning experience of the students.

VIII. CONCLUSION

Although e-learning and online courses are more intensive in the last decade in many countries, it has become mandatory for some countries to maintain the continuation of the learning process in schools, colleges and universities. Within a very short period of time, the researcher's university had to realize the preparation for participation in online lectures and online examinations.

The analogy examined in this study shows that more attempt and effort would make the online learning process more viable for the educational system. The empirical experience gathered in this study has shown the process of increasing feasibility for e-assessment organization.

Furthermore, the results of the analogies were used to establish an initial approach to the implementation of online examinations. The results show that a good guidance of the students through the process, which is unfamiliar to them, is necessary and indispensable to avoid organizational problems. The approach presented here to combine multiple media channels (website and email) to ensure organizational implementation, resulted in no organizational problems in the current study. In addition, it has been shown that the current rudimentary experience of most students with e-exams can be improved by providing self-tests, which are designed to prepare the students with regards to the examination system usage and parallel training of the learning content.

In further studies, it is to be analyzed to what extent the current results can be transferred to other domains. In addition, it should be evaluated whether the process can be simplified without reducing the quality.

REFERENCE

[1] Martin Ebner, Sandra Schön, Clarissa Braun, Markus Ebner, Ypatios Grigoriadis, Maria Haas, Philipp Leitner, and Behnam Taraghi. "COVID-19 epidemic as E-learning boost? Chronological development and effects at an Austrian university against the background of the concept of "E-Learning Readiness"." *Future Internet* 12, no. 6, 2020, pp.-94.

[2] "Policy Brief: Education during COVID-19 and beyond", United Nations, AUGUST,2020.

[3] Kerres, Michael. "Against all odds: Education in Germany coping with Covid-19." *Postdigital Science and Education* 2, no. 3,2020, pp. 690-694.

[4] Terry, Anderson,"The theory and practice of online learning". Athabasca University Press, 2008.

[5] Ulf. Stödberg, "A research review of e-assessment." *Assessment & Evaluation in Higher Education* 37, no. 5 ,2012, pp. 591-604.

[6] "Types of Analogies", June 8.2015.Accessed on : May 5,2021.[Online].Available:https://medium.com/@english_grammar/types-of-analogies-50689fc648e1

[7] Mohamed Ally, "Foundations of educational theory for online learning." *Theory and practice of online learning* 2, 2004, pp. 15-44.

[8] Terry Anderson, "Teaching in an online learning context." *Theory and practice of online learning* 273, 2004.

[9] Mohammed Samaka,and Mohamed Ally. "Model for transforming engineering education using technology-enhanced learning." In *Advances in engineering education in the Middle East and North Africa*,Springer, Cham, pp. 117-138,2016.

[10] Robyn Benson, and Gayani Samarawickrema. "Addressing the context of e-learning: using transactional distance theory to inform design." *Distance Education* 30, no. 1, 2009, pp. 5-21.

[11] Ella R. Kahu, "Framing student engagement in higher education." *Studies in higher education* 38, no. 5 , 2013, pp. 758-773.

[12] Marcia D. Dixson, "Measuring student engagement in the online course: The Online Student Engagement scale (OSE)." *Online Learning* 19, no. 4, 2015, pp. n4.

[13] Heather L O'Brien, Paul Cairns, and Mark Hall. "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form." *International Journal of Human-Computer Studies* 112, 2018, pp. 28-39.

[14] Sung Youl Park, "An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning." *Journal of Educational Technology & Society* 12, no. 3, 2009, pp. 150-162.

[15] Emtinan Alqurashi, "Self-efficacy in online learning environments: A literature review." *Contemporary Issues in Education Research (CIER)* 9, no. 1 ,2016,pp. 45-52.

[16] Lauren Sosniak,A. *Bloom's taxonomy*. Edited by Lorin W. Anderson. Chicago, IL: Univ. Chicago Press, 1994.

[17] Ummay Ubaida Shegupta, René Schmidt, Martin Springwald and Wolfram Hardt. "Audience Response System - an Inclusion of Blended Mentoring Technology in Computer Engineering Education." *IEEE Frontiers in Education Conference (FIE)* ,2020),pp.-1-5.

Lightweight Monocular Depth Estimation on Embedded Systems

Vishnuvarthanan Srinivasan
Department of Embedded Systems
Chemnitz University of Technology
Chemnitz, Germany
shweta.dumbre@s2018.tu-chemnitz.de

Shadi Saleh
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
shadi.saleh@informatik.tu-chemnitz.de

Batbayar Battseren
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
batbayar.battseren@informatik.tu-chemnitz.de

Wolfram Hardt
Department of Computer Engineering
Chemnitz University of Technology
Chemnitz, Germany
wolfram.hardt@informatik.tu-chemnitz.de

***Abstract*— Autonomous vehicles are developed a lot in recent times, which has improved driving and road safety. Various instruments are used to understand the 3D geometric scene around the vehicle. Depth estimation is an important technique to analyze the environment in autonomous driving. Estimating the distance of the object from 2D images using only the cameras is difficult without the depth map of the images. In recent times, convolutional neural networks (CNNs) have developed and used in many futuristic works, including monocular depth estimation (MDE). MDE techniques show a very good estimation of depth from a single RGB image. Using supervised methods require ground truth depth data for training dataset, which are not easy to acquire. Instead, we use only the binocular stereo image pair to train our model and do not require any ground truth depth values. Our training approach helps our CNN to predict depth from a single image with the help of the image reconstruction technique. Only with image reconstruction, we result in poor depth image quality. So we add a left-right disparity consistency check to the image reconstruction, which improves our model performance and robustness. In our training method, we use the KITTI dataset to make our model learn the patterns to predict depth from a single image. Our model produces state-of-the-art results for monocular images in KITTI dataset and outperforms some supervised learning techniques. Even with great results, there are some occlusions occurred in boundaries of the images. MDE can also be used in robotics and virtual reality applications.**

Keywords*— *monocular depth estimation, unsupervised, KITTI, image reconstruction, disparity

I. Introduction

Estimation of depth from single images is easy for a human, but is difficult for computational models with low resource dependent and high accuracy. Depth estimation techniques are used to determine the depth in the image, where many studies have been conducted in various application [1], [2]. There are many approaches for depth estimation in computer vision which depend on structure from motion, shape-from-X, binocular, and multi-view stereo. These approaches depend on the multiple observations of the scene of interest, which can come in the form of multiple viewpoints, or observation of the scene of interest under different lighting conditions. To overcome these restrictions, there is an increase in the number of works that use monocular depth estimation (MDE) as a supervised problem. These learning methods appear to show the statistical relationship between appearances and distances of the objects well. There are many advantages of using MDE which are used in applications like 3D modelling, robotics, autonomous driving, etc. MDE methods work to directly predict the depth of each pixel in an image using the models which have been trained on a large dataset of ground truth depth data. These approaches have been a great success, but have been limited to scenes where a large dataset of image and their corresponding pixel depth are available. The pixel depth in the image is captured by the depth sensors which are LiDAR, RGB-D cameras are used in indoor settings. However, using these sensors have many shortcomings. These sensors have many problems like error and noise characteristics of the particular sensor. In addition to that, the measurements of these sensors are much sparser than the image and do not capture the high detailed depth difference in the image as well. The ground truth can also be generated by the synthetic rendering of the depth maps. The rendered images do not fully replicate the realism of the scene and do not incorporate real image noise characteristics.

Humans are well at performing the MDE by exploiting cues like perspective, scaling relation to the known size of familiar objects, appearance in the form of lighting and shading and occlusion [3]. This combination of both top down and bottom-up cues help in linking the whole scene understanding with our ability to estimate the depth of the object accurately. Although, MDE is an ill-posed problem due to geometric ambiguity where a single 2D image can be produced from an infinite number of distinct 3D scenes. In this work, we take a different technique and overcome the problem of image reconstruction in our model development. Our fully convolutional neural network (CNN) model does not need the depth map of the images, and is instead trained to produce the depth as an intermediate. The model will learn to predict pixel's correspondence between the stereo images, known as camera baseline. There are other methods in solving the MDE problem which has their own limitations. Existing methods are not fully differentiable, making the training suboptimal. When some models training with root mean squared error

(RMSE) they predict reasonably good for a low-resolution depth map, but it is difficult to optimize when we train to predict higher resolution depth maps. We improve on these methods with a new training objective and improvised network architecture that increase the quality of our depth maps. Our model is fast and takes around 35 milliseconds to predict the depth map of the input image of size 512 x 256 on a GPU. The network architecture is an end-to-end unsupervised MDE with an evaluation of several training losses on the KITTI dataset.

Fig. 1. Our depth prediction result on KITTI 2015. At top, single view input image. At bottom, our result of depth map prediction.

II. Literature Review

In this section, we come across past research works done on MDE with several learning-based techniques. Estimating the depth is an important task of understanding the 3D environment of the scene from more than one or more 2D images. Those 2D images can be many overlapping sequences of images captured from different viewpoints, temporal sequences, fixed camera, static environment or change of lighting. Here, we focus on previous research works related to MDE, which has only a single input image and not considering the geometry of the scene or type of objects present.

A. Learning-Based Stereo

Most of the MDE algorithms using the stereo data compute the similarities between every pixel of the first image and every pixel in the second image. Thus, the image pair is rectified, and the disparity can be posed as a 1D search problem for each pixel. The Disparity of the object in the scene is inverse of the depth of the object. Instead of using similarity measures, matching the images as in supervised learning techniques and training a function to predict the correspondence, producing far better results. These binocular correspondence searches are posed as a multi-class classification problem, which has the advantage of a better quality of results and speed. Not only using just the matching function as seen above, Mayer et al. [4] have also used a fully convolutional deep neural network called DispNet that computes the correspondence of the two images. During training, they calculate the disparity of the pixels in the image directly by reducing the regression training loss. DispNet and end-to-end deep optical flow which is their previous network have the same architecture. These methods require a large dataset of stereo image pairs and their accurate ground truth disparity data during training of the model. As we know earlier, this type of dataset is costly and difficult to gather for a real-world environment. So these methods use synthetic data for model training. Synthetic data is more realistic and improving, but still requires the manual creation of new scenes for every different application scenario.

B. Supervised MDE

MDE is to solve the problem where estimation of depth using only a single image during testing. Saxena et al. [5] used a patch-based model known as Make3D. The input image is separated into patches and then estimates the 3D environment and orientation of local planes to show details of each patch. The prediction of the correspondence is made using a linear model trained offline on a large dataset of laser scans that provide ground truth depth maps, and the predictions are combined using Markov Random Field (MRF). The disadvantage of this approach is that this is difficult to model thin structures and the predictions lack in producing real-world outputs. Liu et al. [6] differs from other approaches of using CNN instead of hand-tuning the unary and pairwise terms. Ladicky et al. [7] which is a local approach that uses semantics in their model to make the predictions of pixel depth better. Karsch et al. [8] produced better image-level predictions with whole depth maps from a training dataset that requires the whole training dataset to be available during the testing time, which is a big disadvantage of this approach. Eigen et al. [9], [10] provide the prediction of pixel depth using two scales deep network which is trained on images and their corresponding depth maps. It differs from other approaches with learning from raw pixel values instead of segmentations or hand tunings. Many models have been developed from this approach using methods like conditional random fields (CRFs) to improve accuracy and reduce loss using more robust loss functions. Like stereo-based learning approaches, these approaches require large, high quality, pixel aligned, ground truth depth maps during training time to make a more robust MDE model. Gathering such a dataset is expensive and even with LiDAR data available, calibration of wrong projections requires a manual task and filtering the twist effect created by the orbital nature of the sensor is also required. So, we predict the single image depth estimation by training with the addition of binocular colour image instead of ground truth depth maps from sensors.

C. Unsupervised MDE

There has been development in deep learning models for depth estimation which do not require ground truth depth maps during training time. These approaches face the problem of image reconstruction due to the absence of labelled data. The depth maps are replaced by different cues and losses like image sequences, image synthesis. Flynn et al. [11] has developed a network known as Deep-stereo, which generates different views of an image from pixels in nearby images. In training time, these views are used to predict the appearance of a held-out nearby image. Then the depths are selected from nearby images, based on plane sweep volumes. At testing time, image synthesis is done over small overlapping patches. As many nearby images are required during the testing time, this is difficult to perform. Shadi et al. [12] has introduced colour coded technique to estimate depth from coloured depth maps for the application of collision avoidance in autonomous driving.

Another approach in an unsupervised way is the Garg et al. [13] which uses an encoder-decoder architecture that trained to predict the depth estimation on image alignment loss. The loss amplifies the photometric error of the input image joined with the corresponding other view image using the depth weights generated. The Deep3D network of Xie et al. [14] produces the corresponding right view from an input image, which is a left view in a binocular pair. With image reconstruction, they produce pixel-wise disparity in the image. The result of the right view image pixel is the result from the pixels from the left view image, and the weights are the probability of each pixel. The main drawback of this approach is that the disparity of the image largely consumes the memory of the algorithm, which are difficult when getting a larger output resolution. To solve this problem, our model uses bilinear sampling to generate images and has a training loss.

We have used a CNN model similar to the supervised method DispNet architecture which we have seen above. We treat the MDE as an image reconstruction problem as we can create the disparity without using the labelled data such as ground truth depth maps. By reducing the photometric loss, we can only increase the quality of the image reconstruction, but not the quality of the depth in the image. We implement a left-right consistency check to increase the quality of the generated image depths. The consistency check is used in the post-processing step, which is included in our network.

III. Methodology

In this section, we discuss our single image depth estimation network. We solve the depth estimation as an image reconstruction with an inbuilt left-right consistency network having a training loss, which helps us to train on stereo images without any need of supervision in the form of ground truth depth maps.

A. Depth Estimation as Image Reconstruction

A single image I at testing time, we aim to predict the depth in every pixel of that image with the help of learning from a function *f*, $d' = f(I)$. Most of the current deep learning models have supervised learning method to solve this problem, where they have RGB images and their corresponding ground truth depth maps while training time. It is difficult to gather such ground truth depth maps dataset for a vast kind of scenes. Even with expensive laser scanners, the data we get has effects of natural scenes like movement and reflection. Those data are also needed to be done with calibration of wrong projections requires manual task and filtering the twist effect created by the orbital nature of the sensor. In change, we implement pose depth estimation of the image as an image reconstruction task during training time. Here, we get a calibrated pair of binocular images as input data, then we can train a function that can reconstruct an image from the other and the 3D environment of the scene will get to be learned.

During training time, we get access to the stereo image pair, I^l and I^r, which are left and right RGB images captured at the same moment from two cameras separated by a baseline. We do not directly try to predict the depth in the image, instead, we find the dense correspondence field d^r from the left image which we would be able to reconstruct the right image. The image reconstructed from the left image is denoted as $I^l(d^r)$ as I'^r. Similarly, we will reconstruct the left image from the corresponding right image, $I'^l = I^r(d^l)$. The images are rectified, the disparity of the imaged is the scalar value per pixel of the image that our model will learn to predict. As the distance between the cameras the baseline *b* and the focal length of the camera *f* are given, we can recover the depth *d'* from the disparity predicted from the above factors, $d' = bf/d$.

B. Depth Estimation Network

Our model will predict the depth by the disparities generated that warp the left image of the stereo pair to match the corresponding right image and vice versa [15]. The important advantage of our model is that we can simultaneously match both left-to-right and right-to-left disparities, using just the left image, and get improved depths by adding them to be consistent with one another.

Our neural network produces the predicted images of the left and the right image with the help of a bilinear sampler as backward mapping. This gives us a fully differentiable image generation model. When using the left input image to generate only the disparity of the image, there are depth discontinuities in the generated disparity map. So we eradicate this problem in our network while training, which generates the disparity maps of both the left and right image by sampling from the opposite view images. Even generating the disparity of both view images we require only the left image as an input to our convolutional layers and the right view image from the stereo pair is only used in our training and not in our testing. Using the left-right consistency to generate disparity maps of both images improves our model more accuracy.

Our model with a fully convolutional neural network helps us to train without using ground truth depth maps. Our network consists of two main portions - an encoder and a decoder. The decoder has skip connections with the activation blocks in the encoder, which enables our network to resolve details even in higher resolution. The disparities generated are at four different scales, this factor helps the spatial resolution twice at each subsequent scales. Our network will generate two disparity maps left-to-right and right-to-left even though only the left image is fed as the input to our network.

C. Training Loss

We calculate loss Cs at each scale of output s, concluding the total loss as the sum of losses $C = \sum_{s=1}^{4} C_s$,. Our loss module considers Cs as a combined term of 3 main losses

$$C_s = \partial_{ap}\left(C_{ap}^l + C_{ap}^r\right) + \partial_{ds}\left(C_{ds}^l + C_{ds}^r\right) + \partial_{lr}\left(C_{lr}^l + C_{lr}^r\right) \quad (1)$$

where C_{ap} is the loss that occurred during the reconstruction of the image from the disparity maps of the input image, C_{ds} is the loss that occurred due to the disparity smoothing and C_{lr} is the consistency loss that occurred during the prediction of left and right disparities. Each loss is calculated for both left and right image variants, but only the left image of the stereo pair is given as the input to the convolutional network layers.

Next, we refer to each loss concerning the left image. Similarly, the right image losses require replacing left for right and to sample in the opposite direction.

1) Appearance Matching loss

In training time, the model learns to produce an image by sampling pixels from the opposite image in the stereo pair input. Our image generation model generates the image by using an image sampler from the spatial transformer network (STN) for sampling the opposite image using the disparity map. Then STN uses bilinear sampling, which gives the weighted sum of four input pixels as the output pixel. Indifference from other approaches [13] the bilinear sampler we use in our model is fully differentiable and integrates flawlessly in our fully convolutional neural network architecture. This means we do not need any work to reduce the complexity of the network or the approximation of our cost function.

We combine the L1 and single scale structural similarity index measure (SSIM) in calculating our photometric image reconstruction cost C_{ap}, which is the difference between the input image I^l_{ij} and its corresponding reconstruction I'^l_{ij}, and N is the number of pixels,

$$C^l_{ap} = \frac{1}{N}\sum_{i,j}\alpha\frac{1-SSIM(I^l_{ij}, I'^l_{ij})}{2} + (1-\alpha)\left\|I^l_{ij} - I'^l_{ij}\right\| \quad (2)$$

A simplified SSIM has been used with a 3×3 block filter in replacement of a Gaussian filter and set α=0.85.

2) Disparity Smoothness Loss

We improve the disparities generated by our network and locally smoothing the disparities with an L1 penalty on the disparity gradients ∂d. To overcome the depth discontinuities which occurs at image gradients, we weigh the disparity smoothness cost with the edge-aware term using the image gradients ∂I,

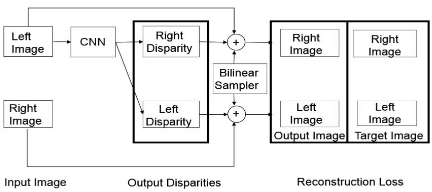
$$C^l_{ds} = \frac{1}{N}\sum_{i,j}\left|\partial_x d^l_{ij}\right| e^{-\left\|\partial_x I^l_{ij}\right\|} + \left|\partial_y d^l_{ij}\right| e^{-\left\|\partial_y I^l_{ij}\right\|} \quad (3)$$

3) Left-Right disparity consistency loss

To generate more accurate and continuous disparity maps, we train our model to produce the disparities of both left and right view images, while only the left-view image is fed as the input into the convolutional layers of the network. To improve the consistency, we place the L1 left-right disparity consistency in our model. This cost is to make the left disparity map equal to the predicted right image disparity map,

$$C^l_{lr} = \frac{1}{N}\sum_{i,j}\left|d^l_{ij} - d^r_{ij+d^l_{ij}}\right| \quad (4)$$

Same as other factors, this loss is equal to the right image disparity map and is calculated at all scales of output.



During testing time, our model generates the disparities at accuracy for the left image dl and the disparities generated has the same resolution as the input left image. Using the defined baseline and the focal length of the camera from the training set, we can convert the generated disparity map to a depth map. The right disparity maps dr are only estimated during the training time and not used in our testing time.

IV. Implementation Details

A. Dataset

Here we perform our model on an unsupervised MDE approach, which we do not need any supervision in the form of ground truth depth data. We cannot use any single image datasets which are absent of the stereo image pairs for training our model. So, we use KITTI 2015 dataset to train our model, where it contains calibrated and rectified stereo image pairs. The dataset is in raw form which consists of 42,382 rectified stereo image pairs from 61 different scenes, with an image size of 1242X375 pixels.

B. Preprocessing

The model is implemented in PyTorch and contains about 31 million trainable parameters and takes around 20 hours to train in a GPU provided in Google Colab Jupyter notebook on a dataset of 20,000 images for 50 epochs. The inference is fast and takes less than 35ms, or more than 28 frames per second for a test image size of 512X256 including the time to transfer data to and from the GPU.

Fig. 3. Sample images from KITTI 2015 dataset. At top, image from left camera. At bottom, image from right camera.

To optimize our model during training, we change the weights of different loss components to $\alpha_{ap} = 1$ and $\alpha_{lr} = 1$. By using scaled sigmoid non-linearity, the disparity of the input images is constrained to be between 0 and d_{max}. The d_{max} is 3X the width of the image at a given output scale. As we use the multiscale output in our model, the disparity of the pixels nearby will be differed by a factor of two in between each output scale because in our decoder we are upsampling twice the output in each scale. To limit this, we scale the smoothing of disparities term α_{ds} with r for equal smoothing factor for each scale at each level. So, $\alpha_{ds}=0.1/r$, where r is the downscaling factor with a specific layer in change to the size of input image pixels that fed into our full convolutional network.

Instead of using rectified linear units (ReLU) for nonlinearity problems in our network, we implemented exponential linear units (ELU). The ReLU prematurely fix the generated disparities of the image at intermediate output scales to a single value when we implemented ReLU in our network, which makes performance improvement difficult. We have used the nearest neighbor upsampling after a convolution, replacing the traditional deconvolution layers in our network. Our model is trained from scratch for 50 epochs, with training images batched in the size of 8 using Adam optimizer [16], where $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the initial learning rate has been set to $\lambda = 10{-}4$ which is followed until the first 30 epochs, and it has been halved every 10 epochs until the end of 50 epochs. At first, we implemented progressive update schedules, which helps us in optimizing the lower resolution image scales in our network first. However, we noticed more stable convergence after optimizing all four output scales at once.

Our data are augmented quickly in the move. The input images are flipped horizontally with a 50% chance, we swap both left and right view of the images, so they are placed in a respective position related to each other. We have also done colour augmentation to our data with a 50% chance, in which we performed random gamma, colour and brightness shift of the image. By sampling, the input images augmentation is done by the uniform values distribution in the range for each shift [0.8,1.2] for gamma, [0.8,1.2] for all 3 colour channels individually and [0.5,2.0] for brightness.

C. Convolutional Neural Network

Resnet 18 is Residual Network-18 which has been used in the implementation. The pre-trained version of resnet 18 which was trained on millions of images from the ImageNet dataset can be used in our network. The CNN architecture we have used is the auto-encoder. The auto-encoder consist of an encoder and a decoder. The resnet 18 network is used as an encoder in the architecture. The input images are batched in size of 8 and of 3 colour channels with the size of (3 x 512 x 256). At first, the images enter a convolutional layer with a filter size of (7 x7). These layers help in the feature extraction in the images. And then images enter a maxpooling 2D layer in which the images are down sampled by reducing their dimensions and assumptions are made about feature containing in the sub regions. Those feature containing sub regions are fed into resnet blocks with convolutional layers and exponential linear unit activation function. The decoder consists of 6 deconvolution layers, which remove the effects of convolution. The feature maps are up-sampled and try to achieve the recreation of the input images. These layers in the encoder and decoder are interconnected, which are known as skips [17]. These skips help us in using higher resolution images in our convolutional layers.

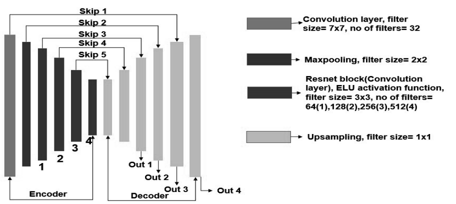


Fig. 4. CNN architecture

D. Post-processing

When using the stereo image pair creates the stereo disocclusions which are disparity ramps on the left side of both the stereo image pair regenerated in our model. To decrease the effects of stereo disocclusions we have implemented a final post-processing process to our output generated. When the test image is given as input image *I*, we augment and horizontally flip the input image I' and then predict the disparity map d'_l. Then we flip back the predicted disparity map, and we get a disparity map d''_l, which are the same as the disparity map generated when the input image is not flipped. This creates the disparity ramps on the right side of the disocclusion and the right side of the input image. We combine both the disparity maps of horizontally flipped and unflipped input images to obtain the final output. The first 5% on the left of the output image is created using, d''_l and the last 5% on the right using d_l. The remaining central part of the output disparity maps is the average of the flipped d'_l and unflipped disparities d_l created. This post-processing on the disparities generated helps us to improve our accuracy in regenerating the left and right images from the final disparities and reduce visual artifacts which help us to cancel the cost of doubling the test image computation time.

V. Result

During testing time, we use only a single image to determine the depth of the pixels in it. We do not feed any additional data such as Velodyne depth data, camera calibration. Our model can predict the depth in the input image for each pixel.

VI. Evaluation

We evaluate our model with the help of 200 high-quality disparity maps provided with part of the training dataset by the official KITTI dataset. These disparity maps are better in quality than the reprojected Velodyne laser depth values. The disparity maps are got from a computer aided design (CAD) model in a moving car. These CAD models get us the less accurate disparity values on transparent things such as car windows and issues at object boundaries. The maximum depth values in the KITTI dataset is 80 meters, and we evaluate our predictions to these values. Results are calculated using the depth metrics and D1-all disparity error from KITTI. These metrics form the measured error in both meters from the ground truth depth maps and depth percentage that are within the threshold from the correct values. In the note, measuring the error in more depth spaces while the ground truth values are given in disparities leads to precision issues. The non-threshold measures can be more responsive to the large depth errors which are occurred due to the prediction errors at smaller disparity values.

For training, K is the KITTI dataset, CS is Cityscapes dataset [18]. Our approach left-right consistency has better results than other two approaches. And our model has almost achieved the results of our state-of-the-art model, despite using less training images.

VII. Limitation

Even with our process of both left-right consistency check and final post-processing technique improves our result quality, we come across the problem of some artifacts seen at occlusion boundaries because the pixels of occlusion regions are not visible in both the stereo images. These problems can be reduced in explicitly reasoning about occlusion in our training [19] [20]. It is also recommended to depend on the camera baseline and the depth sensor, fully supervised learning methods also do not use depth values for all pixels in training. Our approach works on image reconstruction terms, which creates inconsistent depth prediction on transparent surfaces. This issue can be reduced with similarity measures. Our approach requires calibrated and aligned stereo image pairs during our training, which implies that our approach cannot use a single view image dataset in our training. But, it

Fig. 5. Our results after post-processing for a single image

TABLE I. RESULTS ON KITTI 2015

Method	Supervised	Dataset	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Eigen et al. [16]	Yes	K	0.214	1.605	6.56	0.292	0.702	0.89	0.958
Garg et al. [19]	No	K	0.169	1.08	5.104	0.194	0.74	0.89	0.962
LRC	No	K+CS	0.097	0.89	5.093	0.176	0.87	0.96	0.98
Ours	No	K	0.1088	1.039	5.54	0.198	0.859	0.948	0.97

is possible to use our model on a specific application by fine tuning with ground truth depth values.

VIII. CONCLUSION & FUTURE WORKS

We have implemented an unsupervised deep learning network for depth estimation from a single image. We have avoided the usage of ground truth depth maps for the images, which are costly and difficult to collect. Instead, we use binocular stereo image pairs from the KITTI dataset. Our novel loss function is calculated from the consistency in our predicted depth maps from each image view during our training time, which improves our depth predictions. Our model can be seen providing good performance over unseen datasets and predicts visually plausible depth maps.

In Future, we would like to improve our current depth predictions for every independent frame and incline our model to perform depth prediction on videos. It would be interesting if our model can perform and reduce occlusion effects in the full occupancy depth prediction of the scene.

IX. ACKNOWLEDGMENT

I am thankful to Prof. Dr. Dr. h. c. Wolfram Hardt, Chair of Computer Engineering, TU Chemnitz, for providing the opportunity to perform the Research internship and a platform to learn about the latest technology used in automotive. I would also like to show my gratitude to the faculties Shadi Saleh, Batbayar Battseren for sharing their pearls of wisdom with me during the Internship. I also thank the faculties and students of University of Jan Evangelista in Ústí nad Labem for sharing their knowledge and experience during IKON project. Last but not the least I would like to thank all the anonymous persons who have done work in the field of Depth Estimation.

REFERENCES

[1] S. Saleh and S. A. Khwandah, "Traffic Signs Recognition and Distance Estimation using a Monocular Camera," 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 407–418, 2019. [Online]. Available: http://ceur-ws.org/Vol2514/paper104.pdf

[2] S. Shadi, S. Hadi, and M. A. Nazari, "Outdoor Navigation for Visually Impaired based on Deep Learning," 6th International Conference Actual Problems of System and Software Engineering, vol. 2514, pp. 397–406, 2019. [Online]. Available: http://ceur-ws.org/Vol-2514/paper102.pdf

[3] I. P. Howard and B. J. Rogers, Perceiving in depth, ser. Oxford psychology series. New York: Oxford University Press, 2012, no. 29.

[4] N. Mayer and E. Ilg, "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation," IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016. [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2016/MIFDB16

[5] A. Saxena, Min Sun, and A. Ng, "Make3D: Learning 3D Scene Structure from a Single Still Image," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 824–840, May 2009. [Online]. Available: http://ieeexplore.ieee.org/document/4531745/

[6] F. Liu and C. Shen, "Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields," 2015. [Online]. Available: https://arxiv.org/pdf/1502.07411.pdf

[7] L. Ladicky, J. Shi, and M. Pollefeys, "Pulling Things out of Perspective," in 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, Jun. 2014, pp. 89–96. [Online]. Available: https://ieeexplore.ieee.org/document/6909413

[8] K. Karsch, C. Liu, and S. B. Kang, "Depth Transfer: Depth Extraction from Video Using Non-Parametric Sampling," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2144–2158, Nov. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6787109/

[9] D. Eigen and C. Puhrsch, "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network," 2014. [Online]. Available: https://papers.nips.cc/paper/2014/file/7bccfde7714a1ebadf06c5f4cea752c1Paper.pdf

[10] D. Eigen and R. Fergus, "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture." [Online]. Available: https://arxiv.org/pdf/1411.4734.pdf

[11] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "DeepStereo: Learning to Predict New Views from the World's Imagery," arXiv:1506.06825 [cs], Jun. 2015, arXiv: 1506.06825. [Online]. Available: http://arxiv.org/abs/1506.06825

[12] S. Saleh, S. Manoharan, J. Nine, and W. Hardt, "Towards robust perception depth information for collision avoidance," in 2020 IEEE Congreso Bienal de Argentina (ARGENCON), 2020, pp. 1–4.

[13] R. Garg, V. K. BG, G. Carneiro, and I. Reid, "Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue," arXiv:1603.04992 [cs], Jul. 2016, arXiv: 1603.04992. [Online]. Available: http://arxiv.org/abs/1603.04992

[14] J. Xie, R. Girshick, and A. Farhadi, "Deep3D: Fully Automatic 2Dto-3D Video Conversion with Deep Convolutional Neural Networks,"

arXiv:1604.03650 [cs], Apr. 2016, arXiv: 1604.03650. [Online]. Available: http://arxiv.org/abs/1604.03650

[15] S. Saleh, S. Manoharan, and W. Hardt, "Real-time 3D Perception of Scene with Monocular Camera," Embedded Selforganizing Systems, vol. 7, no. 2, pp. 4–7, 2020.

[16] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs], Jan. 2017, arXiv: 1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980

[17] S. Saleh, S. Manoharan, J. Nine, and W. Hardt, "Towards robust perception depth information for collision avoidance," in 2020 IEEE Congreso Bienal de Argentina (ARGENCON), 2020, pp. 1–1.

[18] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised Monocular Depth Estimation with Left-Right Consistency," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, Jul. 2017, pp. 6602–6611. [Online]. Available: http://ieeexplore.ieee.org/document/8100182/

[19] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering Occlusion Boundaries from an Image," International Journal of Computer Vision, vol. 91, no. 3, pp. 328–346, Feb. 2011. [Online]. Available: http://link.springer.com/10.1007/s11263-010-0400-4

[20] A. Humayun, O. Mac Aodha, and G. J. Brostow, "Learning to find occlusion regions," in CVPR 2011. Providence, RI: IEEE, Jun. 2011, pp. 2161–2168. [Online]. Available: https://ieeexplore.ieee.org/document/5995517/

[21] H. Fu, M. Gong, C. Wang, and D. Tao, "A Compromise Principle in Deep Monocular Depth Estimation," arXiv:1708.08267 [cs], Jun. 2018, arXiv: 1708.08267. [Online]. Available: http://arxiv.org/abs/1708.08267

[22] Y. Kuznietsov, J. Stuckler, and B. Leibe, "Semi-Supervised Deep¨ Learning for Monocular Depth Map Prediction," arXiv:1702.02706 [cs], May 2017, arXiv: 1702.02706. [Online]. Available: http://arxiv.org/abs/1702.02706

[23] J. Hu, Y. Zhang, and T. Okatani, "Visualization of Convolutional Neural Networks for Monocular Depth Estimation," arXiv:1904.03380 [cs], Apr. 2019, arXiv: 1904.03380. [Online]. Available: http://arxiv.org/abs/1904.03380

[24] A. Bhoi, "Monocular Depth Estimation: A Survey," arXiv:1901.09402 [cs], Jan. 2019, arXiv: 1901.09402. [Online]. Available: http://arxiv.org/abs/1901.09402

[25] S. Elkerdawy, H. Zhang, and N. Ray, "Lightweight Monocular Depth Estimation Model by Joint End-to-End Filter Pruning," in 2019 IEEE International Conference on Image Processing (ICIP). Taipei, Taiwan: IEEE, Sep. 2019, pp. 4290–4294. [Online]. Available: https://ieeexplore.ieee.org/document/8803544/

[26] V. Guizilini, R. Ambrus, S. Pillai, A. Raventos, and A. Gaidon, "3D Packing for Self-Supervised Monocular Depth Estimation," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, Jun. 2020, pp. 2482–2491. [Online]. Available: https://ieeexplore.ieee.org/document/9156708/

[27] A. Mallya and S. Lazebnik, "PackNet: Adding Multiple Tasks to a Single Network by Iterative Pruning," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, Jun. 2018, pp. 7765–7773. [Online]. Available: https://ieeexplore.ieee.org/document/8578908/

[28] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep Ordinal Regression Network for Monocular Depth Estimation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, Jun. 2018, pp. 2002–2011. [Online]. Available: https://ieeexplore.ieee.org/document/8578312/

Training of the Masters' Students in the Smart City Program Using the Internet of Things Course

Zhanar Akhayeva
Department of Computer Science
L.N. Gumilyov Eurasian National University
Nur-Sultan city, Kazakhstan
ahaeva07@mail.ru

Alma Zakirova
Department of Computer Science
L.N. Gumilyov Eurasian National University
Nur-Sultan city, Kazakhstan
alma_zakirova@mail.ru

Aigul Nurpeissova
Department of Foreign languages
L.N. Gumilyov Eurasian National University
Nur-Sultan city, Kazakhstan
ai.nurpeissova@gmail.com

Abstract — **This work is devoted to the SMARTCITY project: Innovative Approach Towards a Master Program on Smart Cities Technologies, the use of the Internet of Things (IoT) course in Smart City training programs for undergraduates. The purpose of such a training program is to consider issues related to information and communication technologies used in Smart City projects. The course used in this work aims to introduce students to modern information technologies and the use of IoT used in Smart City.**

Keywords— Internet of Things, Smart City, educational course, W3C, OGC.

Introduction

Currently, issues related to the Internet of Things attract a lot of attention, and since IoT is directly applied in Smart City, therefore we are considering this direction more deeply. However, most of the descriptions and reviews are frankly journalistic in nature, focusing on descriptions of processes and their impact on everyday life. At the same time, all these abbreviations hide quite real standards, applications and methods of information technologies. It is these aspects that have been actively standardized recently. Accordingly, it would be important to consider the technical aspects of these areas. Every year, the global demand for IoT developers is increasing. This, of course, brings the issues of training to the fore [1].

IoT technologies did not arise from scratch, it has its own direction in the disciplines related to information and computer technologies. At the same time, its development is associated with certain specifics, covers quite specific sections in programming, architecture of computer networks and applications, etc.

In the current state of the project, we are talking about a six-month semester course, which aims to introduce students to modern information technologies behind such a direction as IoT. In this case, we mean students studying in areas related to Computer Science. For example, the subject of such a course corresponds to such a master's program at the L. N. Gumilyov ENU as "Smart City" (in the direction of Computer Science). How is this direction related to Smart City, IoT or programming of embedded systems, i.e. it is the main part of Smart City.

II. Standards Applied In Smart City

This section discusses the general principles of working on standards. Therefore, we will begin our consideration with several topics of standardization that are relevant to the topic under consideration.

The World Wide Web Consortium (W3C) is an organization that develops and implements technological standards for the World Wide Web. The consortium is headed by Sir Timothy John Berners-Lee, the author of many developments in the field of information technology [2]. The Open Geospatial Consortium (OGC) is an international organization for the development of standards in the field of geoinformation services. Until September 2004, the consortium was called the Open GIS Consortium, founded in 1994 [3]. Digital Imaging and Communications in Medicine (DICOM) is an industry standard for creating, storing, transmitting and visualizing medical images and documents of examined patients [4]. The Clinical Data Interchange Standards Consortium (CDISC) is an organization that develops standards in the field of medical information. The goals are quite clear – an electronic patient card, which, if necessary, can be read in any medical institution in the world [5].

Organization for the Advancement of Structured Information Standards (OASIS) is a global consortium that manages the development, convergence and adoption of industrial e — commerce standards within the international information community. This consortium is the leader in the number of issued standards related to Web services. In addition, he is engaged in standardization in the field of security, e-commerce; the public sector and the markets of highly specialized products are also affected. OASIS has over 5,000 participants representing more than 600 different organizations from 100 countries around the world [6].

Object Management Group (OMG) is a consortium engaged in the development and promotion of object — oriented technologies and standards. This is a non-profit association that develops standards for creating interoperable, that is, platform-independent, applications at the enterprise level. About 800 organizations — the largest software manufacturers-cooperate with the consortium. [7]

Let's start with the last two. The narrower DICOM is actually adopted as an international standard (ISO standard 12052:2006 "Health informatics -- Digital imaging and communication in medicine (DICOM) including workflow and data management") by manufacturers of radiology devices (including tomographs and nuclear magnetic resonators), and by practitioners. CDISC, which started its work in the USA since 2000, has been working in the European Union, since 2001 in Japan, since 2008 in China, since 2009 in South Korea. CDISC has long been accredited by ISO and with its participation, international standards are being prepared, which are based on CDISC standards.

A similar story is with the W3C and OGC – they have long had agreements not only with ISO, but also with IEEE, IEC, and a number of other international standardization organizations, and their standards are becoming the basis of international ones. The joint work of OGC with the European trendsetter on three – dimensional models-Inspire is fundamentally important [8]. The models are designed in the form of detailed legal requirements of the European Community. No one can design, build and operate almost any facility in Europe without taking into account these requirements. It is also important that all these organizations are linked by working agreements and, if necessary, can concentrate their resources on the chosen direction.

The "Internet of Things" (IoT) market of Kazakhstan is to some extent the heir of M2M technological solutions, which have been actively developing in the Republic for more than ten years. IoT is not only the transfer of data between the active elements of the inter-machine network, but also a complex of vertically integrated industry solutions, including sensors, data processing platforms from them, software and actuators. As part of the study of the IoT market of Kazakhstan, a PEST analysis of the market was conducted to assess the qualitative factors of market development, the results of which are shown in the Table I below. [9]

TABLE I

Factors	Relevance	Impact
POLITICAL FACTORS (P)		
The legislation in the IoT field has been actively developing since 2017. The new standards for the implementation of IoT solutions in various sectors of the economy are being introduced	Average	Positive
The Government of Kazakhstan actively promotes the introduction of IoT projects in the economy by developing and financing programs for the development of the following segment (financing of IoT projects from the Unified State Fund, "Digital Kazakhstan" program)	High	Positive
THE IMPACT OF THE ECONOMY (E)		Positive
The growth of the country's gross domestic product as a whole has a favorable effect on investment in new IT projects	Average	Positive
In the energy sector of Kazakhstan, the issue of reducing electricity costs is acute. The need to optimize the work of segments of the energy industry leads to more active support for IoT projects in this segment, both from the state and from companies.	Average	Positive
SOCIAL FACTORS (S)		Positive
Low cost of mobile communication services in the Republic of Kazakhstan	Low	Positive
The consumption of mobile and fixed-line services continues to grow. Broadband access and mobile communications are expanding penetration outside of major cities.	Average	Positive
There are preconceived concerns on the part of users of IT products about the potential replacement of large IT solutions by the workforce.	Low	Positive
TECHNOLOGICAL INNOVATIONS (T)		Positive
The availability of LoRaWAN technology and the relatively low cost of deploying a network using this technology contribute to the active development of IoT projects in the cities of the Republic of Kazakhstan, especially in the areas of "smart" cities and smart housing and communal services	High	Positive

[a]J'son & Partners Consulting

The results of the analysis show that a positive climate has developed in the Republic of Kazakhstan for the development of the "Internet of Things" market, it is worth noting the purposeful efforts of the state in the development of the industry. Back in 2015-2016, 9 national standards in the field of the "Internet of Things" were adopted, which laid down the policy and basic principles of industry development in the Republic of Kazakhstan (RK). In addition, considerable attention is paid to the "Internet of Things "in the" Digital Kazakhstan "program, in addition, state programs for the elimination of "digital inequality" indirectly contribute to the development of the IoT industry.

In fact, the purposeful efforts of the state are the main driver of the development of the IoT market in the Republic of Kazakhstan. Together with the relatively high growth rates of the economy of the Republic of Kazakhstan, they, according to experts of J'son & Partners Consulting, will ensure high growth rates of the market. [9]

n 2014, the above-mentioned BSI adopted the world's first Smart City standard (PAS 181) [10], and the following year it was supplemented and developed by a number of additional standards. A little later, but also in 2014, ISO released its Smart City 37120 standard [11], but these are different standards! In fact, this is a whole set of standards. The ISO standard tells how to measure the level of urban service and the quality of life of citizens, and in the same year 2014 it became the UK standard BS ISO 37120: 2014, becoming one of the current standards for supporting a Smart city. Actually, PAS 181 is intended for practical planning of the development of cities in the UK, and along with other standards on this topic, it became the basis for the rapid development of this direction in the UK, supplemented at the same time by the powerful organizational support of the Government called CATAPULT [12]. Naturally, British specialists in W3C, OGC, DICOM, CDISC were also actively involved. For example, OGS has introduced a GIS data model of Smart cities and standards for geo-sensors of the Internet of Things, and CDISC and DICOM are actively involved in the construction of digital medicine in the UK (by the way, there are already more than 80% of digital hospital cards in the UK, and a little more than 25% in the USA).

Spanish standardization specialists also adopted the ISO standard under the number UNE-ISO 37120 and supplemented it with their two standards UNE 178301 and UNE 178303. However, these are standards on the topic of urban asset management. 75 cities of the world have joined the Open & Agile Smart Cities initiative, which is designed to standardize the work with data in smart cities [13].

The reason for this is that the situation with the rules and regulations is different in different countries and if the British can rely on the success of implementing information modeling or BIM systems and, for example, on the standard for underground infrastructure or wireless sensor networks, and appropriate application practices and personnel, then the situation, for example, in Spain is completely different. In addition, there is not only cooperation between the Smart Cities of the world, but also fierce competition for digital economy companies, investors, tourists, etc. China has acted even more pragmatically in this matter, ordering the localization of British standards for Smart City in combination with other standards they need from BSI.

The situation with the Internet of Things is even more complicated. International standardization organizations cannot work in this direction, since there is not a single national standard for IoT, moreover, in April 2015, NIST officially stated that IoT is not defined. This is due to the extremely rapid development of applications: medicine, energy, transport, etc. IoT is already used in various spheres of human activity. These applications are very diverse-from marine platforms to medical applications and aviation. Actually, from the broad interpretation of the use of IoT terminology from application development to actually physically working sensors, huge difficulties in setting standards follow. In May 2015, the IEEE tried to release a material "On the question of defining the Internet of Things", but it came down to presenting the positions of various respected organizations [14]. Thus, the W3C considers IoT from the point of view of the Web of Things [15], and NIST in the context of cyber-physical systems [16]. The latter position is closest to those who are engaged in production facilities and, in particular, German specialists in Industry 4.0 [17].

EXISTING TRAINING PROGRAMS

This section contains training programs in the direction of Smart City.

a) Master's program in Smart Cities and Urban analytics [18]. Covers the following issues: networks and communications, traffic flow planning, real-time systems, geo-information systems, modeling systems. The main focus is urban planning and management. This is something that applies specifically to Smart Cities.

b) ICT & Next generation networks [19]. It covers the following issues: NGN, broadband access, QoS. Optical and wireless networks, telecommunications network architecture, 4G, mesh networks. The main focus, as can be seen from the content – is network interaction

c) Master of City Science – Internet of Things [20]. Current projects (status) Internet of Things for urban and transport tasks.

d) Master of City Science – service applications [21]. Main directions: applications for urban services, open data, application economics.

e) There is an ICT Labs Embedded Systems program for network technologies of interest [22], with a specialization in Internet of Things.

f) SAP offers an online course on the Internet of Things [23]. Materials of a fairly general nature are just an introduction to the problem.

g) Other online courses include the Open University IoT program [24]. It uses IoT elements for general computer science education. In particular, an Arduino-based board is used for programming training.

h) The University of Oxford offers the course Data Science for the Internet of Things [25]. The questions of the application (use) of machine learning for IoT (measurement analysis) are considered.

i) Coursera [26] offers a specialization (a set of courses, a diploma project and a certificate) on the Internet of Things.

j) The Professional Training Program at MIT also offers an IoT training program [27]. This program includes the following sections [28]: IoT architecture, sensor data processing, SLAM, autonomous devices (cars, robots), IoT standards, wearables, security, Web of Things, wireless protocols, data storage and analysis, security, human-machine interfaces. The module of the course on application development includes smart homes, smart cities, smart materials, medical applications and cyber-physical systems. Given the stellar composition of lecturers at MTI, this course looks the most advanced. By the way, the relationship between IoT and Smart Cities follows from this structure. According to MIT, Smart Cities has an IoT application. That is, IoT is just a base for Smart Cities, Smart Home, etc.

k) Kings College London offers a free overview course on IoT [29]. The University of Washington offers a practical course on IoT (Raspberry Pi and other devices) [30].

l) HP also has its own training course on IoT [31]. This is a two-day course, although, formally, there are quite a lot of sections. Among the declared topics: RFID, WSN, Smart applications (Smart Cities, Smart Home, Smart Metering, Smart Health). According to the general characteristics – an overview course for managers.

m) Waterford Institute of Technology (Ireland) offers a bachelor's degree program in the field of Internet of Things [32].

n) Intel offers an open course called Internet of Things [33], but, in fact, promotes programming for its own Edison platform.

IOT COURSE

The Smart City educational program for the project "SMARTCITY: Innovative Approach Towards a Master Program on Smart Cities Technologies" is dedicated to information technologies and their processes and applications for interaction with various measuring devices (sensors). It consists of combining four main and very important components: application architecture, network interaction standards, data storage systems and the principles of their analysis and processing. Accordingly, the program includes four main sections.

The first of them discusses modern standards in the field of M2M and IoT. IoT-GSI, OpenIoT, FI-WARE, OMA, etc. – the number of proposed solutions, in varying degrees of readiness, is quite large. At the same time, they can promote different approaches to building systems, collecting and processing data. This is, in fact, a consideration of possible architectural solutions for building IoT systems. Undergraduates will have to get an idea of existing architectural solutions, their comparative characteristics and applicability depending on external conditions.

The second section will be devoted to network solutions for IoT. Here, in particular, it is planned to consider such solutions as 802.15.4, 6LoWPAN, COAP, RPL. It is closely related to the previous section, since the choice of network protocols can determine the architecture of applications.

Interaction with a large set of external devices (sensors) will most often be associated with the collection and storage of some information (measurements). Accordingly, the tasks associated with storing and processing large data sets are among the main ones for IoT applications. The third section of the course deals with issues related to the storage of large data sets. The specifics of the IoT lead to the fact that the main attention will be paid to the processing of data streams (stream data processing) and systems that support it (for example, Apache Storm [34] and Apache Spark [35]).

The last section is devoted to the consideration of real-time data processing methods: classification and determination of anomalies. At the same time, the features of the IoT subject area require consideration, first of all, of streaming algorithms.

Functions and types of professional activity:

- development, integration and operation of information systems of the Internet of Things class; connection of sensors and devices to cloud platforms, data visualization from real equipment for various use cases of IoT systems, transfer of processed data to external systems; development of Web applications and mobile applications; ensuring information security of the system, storing big data of information.
- planning, forecasting and management of IT projects, business processes. Development of algorithms for solving control problems of automation objects.
- collection and processing of statistical data, monitoring of information systems; conducting data mining based on mathematical methods and data representation models to solve key business problems and conducting research in the field of application and development of Smart City technologies.

Building competencies in the course:

a) understanding the architecture of IoT and M2M applications;
b) models of network interaction in IoT;
c) understanding the models of network interaction in IoT;
d) orientation in the network standards used in IoT;
e) understanding the data models used in IoT applications;
f) the ability to choose data models depending on the requirements of orientation in data processing methods.

CONCLUSION

Preparation of masters of IT education on the basis of a multidisciplinary approach, possessing fundamental knowledge and practical skills in the field of analysis for the development, management of Internet systems and services, and IoT.

The uniqueness of the innovative educational program lies in the formation of comprehensive knowledge and skills necessary for training personnel in the field of SmartCity. The results of the training are aimed at forming the qualifications of masters who have modern IT in training, business analysis methods and information resource management, to carry out practical activities based on scientific activities on SmartCity technologies [36].

REFERENCES

[1] Намиот Д. Е. " Об учебных программах по Internet of Things" //International Journal of Open Information Technologies, vol. 5., 2015, pp.35-38

[2] OGC. (2008, Dec.19). Open Geospatial Consortium [Online]. Available: http://www.opengeospatial.org/

[3] DICOM. (1998). Digital Imaging and Communications in Medicine [Online]. Available: http://dicom.nema.org/

[4] CDISC. (1997). Clinical Data Interchange Standards Consortium [Online]. Available: http://www.cdisc.org/

[5] OASIS. (1993). Organization for the Advancement of Structured Information Standards [Online]. Available: https://www.oasis-open.org/

[6] OMG. (1989). Object Management Group [Online]. Available: https://www.omg.org/

[7] OGC. (2008). Open Standards and INSPIRE - OGC Portal [Online]. Available: http://www.opengeospatial.org/pressroom/marketreport/inspire

[8] TAdviser. (2005). Technologies and Adviser [Online]. Available: https://www.tadviser.ru/index.php

[9] BSI. (2004) PAS 181 Smart city framework [Online]. Available: http://www.bsigroup.com/en- GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-181- smart-cities-framework/

[10] Smart City. (2007) Smart Cities Activator 37120 [Online]. Available: http://smartcitiescouncil.com/article/dissecting-iso- 37120-why-new-smart-city-standard-good-news-cities

[11] Catapult. (2010). Catapult Programme [Online]. Available: https://www.catapult.org.uk/

[12] OASC. (2015, Mar.). Open & Agile Smart Cities [Online]. Available: http://www.oascities.org/

[13] IEEE IoT. (2015, May 27). Towards a Definition of the Internet of Things (IoT) [Online]. Available: http://iot.ieee.org/definition.html

[14] Guinard D., Trifa V. Towards the web of things: Web mashups for embedded devices //Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web (MEM 2009), in proceedings of WWW (International World Wide Web Conferences), Madrid, Spain. – 2009. – pp. 15.

[15] Baheti R., Gill H. Cyber-physical systems //The impact of control technology. vol. 12, 2011, pp. 161-166.

[16] Heng S. Industry 4.0: Upgrading of Germany's Industrial Capabilities on the Horizon //Available at SSRN 2656608., 2014.

[17] UCL. (1826). Smart Cities and Urban. Analytics [Online]. Available: http://www.bartlett.ucl.ac.uk/casa/programmes/ postgraduate/msc- smart-cities-and-urban-analytics.

[18] UPM. (1971). ICT&Next generation networks [Online]. Available: http://www.citysciences.com/specialties/ict-next-generation-networks/

[19] UPM. (1971). Master in City Science Internet of Things http://www.citysciences.com/specialties/internet-of-things/

[20] UPM. (1971). Master in City Science – Service Applications http://www.citysciences.com/specialties/services-applications/

[21] ICT. (1989). Labs Embedded Systems [Online]. Available: http://www.masterschool.eitictlabs.eu/programmes/es/

[22] Open SAP. (2013). Internet of Things [Online]. Available: https://open.sap.com/courses/iot1

[23] Gerd Kortuem. Open University Internet of Things (IEEE Computer, Feb. 2013 (vol. 46 no. 2), pp. 53-61.). [Online]. Available: http://www.kortuem.com/internet- of-things-education/

[24] The Open University. (2014). My Digital Life [Online]. Available: http://www.open.ac.uk/courses/modules/tu100#details

[25] The Open University. (2014). Coursera IoT [Online]. Available: https://ru.coursera.org/specializations/iot

[26] MIT. (2002). IoT [Online]. Available: https://professional.mit.edu/

[27] MIT. (2002). Internet of Things: Roadmap to a Connected World [Online]. Available: https://mitprofessionalx.mit.edu/courses/course-v1:MITProfessionalX+IOTx+2016_T1/about

[28] FutureLearn. (2012, Dec.). Kings College London [Online]. Available: https://www.futurelearn.com/courses/internet- of-things

[29] UW. (1861). University of Washington [Online]. Available: http://www.pce.uw.edu/certificates/internet-of-things.html

[30] HP. (1939). Internet of Things [Online]. Available: http://h20195.www2.hp.com/v2/GetPDF.aspx/c04656912.pdf

[31] WIT. (1970). Internet of Things [Online]. Available: https://www.wit.ie/courses/type/science/department_of_computing_maths_physics/bsc-hons-in-the-internet-of-things#tab=description

[32] Intel Software Academic Program. (2015). Intel IoT [Online]. Available: http://intel-software-academic- program.com/pages/courses#diy

[33] Intel Software Academic Program. (2015). Intel IoT Course: [Online]. Available: https://github.com/guermonprez/intel-academic-IoT-course

[34] Jain A., Nalya A. Learning Storm. – Packt Publ., 2014. ISBN 9781783981335

[35] Riedel M. et al. Smart Data Analytics Methods for Remote Sensing Applications. – 2014.

[36] ENU. (2021). L.N. Gumilyov Eurasian National University [Online]. Available: https://asf.enu.kz/storage.pdf