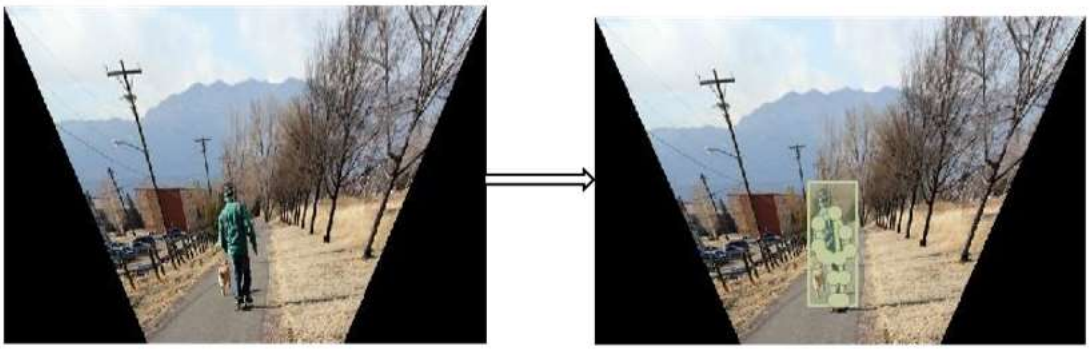
Abstract

An omnidirectional camera is utilized to detect the activity of elderly people living by their own. The first step for achieving this goal is to detect the position of the key points (joint of the human body) in the omnidirectional camera images. Under these circumstances, a human body estimator was implemented using Neural networks and deep learning technology.

In this research work, two different versions with different number of stages (6-stages, 8-stages) of OpenPose, a human body estimator, which detects multiple people in an image, has been used. The original pose estimator was trained to detect people in normal images taken in normal perspective. To achieve the purpose of this research the pose estimator was retrained using transformed images to new perspective similar to omnidirectional perspective. The preprocessing of the training data is illustrated in figure 1.

a) loading keypoints to normal image using COCO API



b) loading transformed keypoints on transformed image using COCO API

Figure 1 Training data pre-processing

A new dataset for training the pose estimation was created using the MSCOCO dataset, people images where extracted from this dataset, then a four-point transformation was implemented to the extracted images, and their annotations. As a result, a new transformed dataset was created for this purpose. To increase the accuracy of the detection, a deeper version of the OpenPose neural network was created. An evaluation omnidirectional dataset, which contains annotated omnidirectional images for single and multiple persons, was created to test and evaluate the trained models.

The trained models were tested on the created omnidirectional dataset. Both models were able to detect human position in omnidirectional images. Results showed that the 8-stages model has better accuracy than the 6-stages model, while the 6-stages model has smaller inference time than the 8-stages model. The output of the trained models are illustrated in figure 2.

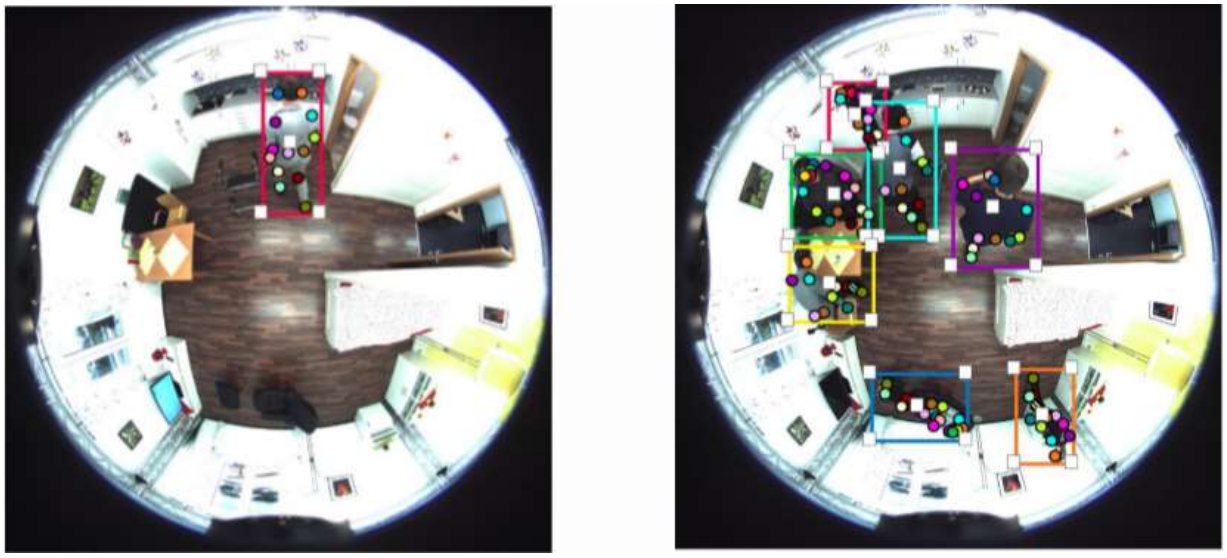


Figure 2 Detection results of trained models on single-person and multi-person images