Abstract

This thesis presented YOLO V3 and SSD Resnet 50 networks for detecting human poses “standing”, “sitting”, “lying” and rollator in omnidirectional camera images.

There are two approaches. One of them is named SSD Resnet 50: single shot multibox detector with ResNet 50 as feature extractor. The other one is named YOLO V3: You only look once version 3. These two methods go from input image to a tensor of scores with one big convolutional network and are simpler and faster than the method of R-CNN.

A new dataset with training and evaluation samples had to be created for the training of the networks. A collection of rollator images are downloaded from the internet. The images are selected to simulate the overlooked viewpoint of omnidirectional images. The images are then manually annotated. The PIROPO dataset and BOMNI dataset had been modified to form the training dataset for the poses classes. A dataset of DST lab images is annotated for the evaluation. Tables 1 and 2 show the statistics of the datasets.

Training category	Numbers	Percentage
Standing	10033	78.5%
Sitting	1798	14.1%
Lying	364	2.8%
Rollator	589	4.6%
All	12784	100%

Table 1: Statistics of different poses categories in training dataset

Class	Number of ground truth objects per class
Lying	33
Sitting	32
Standing	372
Rollator	100

Table 2: Statistics of different poses categories in evaluation dataset

In order to get a competitive accurate and faster result, these two methods were delved into and compared in terms of average precision for different training categories. This research first trained the rollator detection using YOLO V3, while the result is not as accurate as the thought at the beginning of this research. Then we modified the original rollator images and utilized YOLO V3 again to train “standing”, “sitting”, “lying” and rollator. Afterwards this thesis turned to training SSD Resnet50 network for all the four categories. For 480 × 480 input, YOLO V3 achieves 16.45% mAP on chosen test dataset. Compared to SSD Resnet 50 that has better accuracy of 27.5% mAP with a bigger input image size of 640 × 640.