Preliminary Study on Detection of White-Cane Users by Surveillance Cameras and YOLO

Qixi He^{a,*}, Hotaka Takizawa^b, Akihisa Ohya^b, Makoto Kobayashi^c, Mayumi Aoyagi^d

^aMaster's Program in Computer Science, University of Tsukuba, 300-2622, Japan

^bFaculty of Engineering, Information and Systems, University of Tsukuba, 305-8573, Japan

^cFaculty of Health Sciences, Tsukuba University of Technology, 305-8573, Japan

^dDepartment of Special Needs Education, Aichi University of Education, 448-8542, Japan

*Corresponding Author: he.qixi@pr.cs.tsukuba.ac.jp

Abstract

The problem of personal accidents of visually impaired individuals in public areas has become a social issue. In this study, we propose a detection method of white-cane users, which are visually impaired individuals, based on surveillance cameras and YOLO. The proposed method was applied to actual videos, and several experimental results were shown.

Keywords: Visually impaired individuals, White-canes, YOLO, Surveillance cameras

1. Introduction

There are more than 310,000 visually impaired people in Japan (1). Most of them use rail transportation in their daily life, but many station platforms are very dangerous for them. According to Japanese Ministry of Land, Infrastructure, Transport and Tourism, there have been more than 700 falling accidents in station platforms since 2010, with 12 deaths after 2016 alone, nine of which occurred at stations with fewer than 100,000 people (2).

The reasons for such results can be considered as two main reasons. Firstly, the current installation rate of platform gates is only 9% and its priority is given to stations with more

than 100,000 people (3). The second is that no station personnel were present at the time of all the accidents. At this point, it will take a lot of time for platform gates to become widespread and to be effective in reducing accidents. It is necessary to increase the number of station staffs to assist the visually impaired.

However, it is difficult to accurately grasp the existence of visually impaired individuals only relying on the observation of station staffs. Therefore, this study proposes a detection method of visually impaired individuals based on surveillance cameras. In public areas, such as train stations, visually impaired individuals are required to walk with white canes that have distinct features of notifying other people of the existence of the visually impaired. In this paper, we focus on the detection of white canes from the video frames obtained by surveillance cameras by use of YOLO.

2. Data Set and Data Augmentation

We have prepared a data set of 258 images, of which 208 are for training and 50 are for validation. The resolution of these images is 640×640 pixels. We use `labelImg(5)` to label white canes, and save the four vertices of rectangles including the white-canes as shown in Figure 1.

Fig1. Labeling of a white-cane

In order to improve the detection performance of white canes under different backgrounds, lighting conditions and shooting distances, images were preprocessed by horizontal flip, random displacement, random scaling, random HSV changing, random rotation, mosaic (6) and mixup (7) operations as shown in Figure 2 to 8 to enhance the diversity of the training set.

The mosaic operation combines several small-scale images into one image, which can effectively improve training efficiency.

The mixup operation superimposes two images after changing their transparency to form a new image. It can increase robustness and improve performance in unlearned scenarios.

By randomly combining these preprocessing methods, we expanded the training set to 624 images.

Fig2. Horizontal flip

Fig 3. Random displacement

Fig 4. Random scaling

Fig 5. Random HSV

Fig6. Random rotation

Fig7.Mosaic

Fig8. Mixup

3. YOLO

YOLO (4) is an object detection algorithm based on neural network model. In this study, we use the YOLOv5 (8) model that consists of four layers (i.e. input, backbone, head, and output) as shown in Figure 9.

Fig9. YOLOv5 structure

The input layer accepts images and divides them into three RGB channels. The backbone layer is composed of three modules: Focus (9), C3, and CSP (10), and is implemented using Pytorch. The main function of the backbone is to extract features. The head layer is composed of PANet (11) and Detect. The PANet module can detect the white cane in three scales of 1/8, 1/16, and 1/32, and the Detect module can obtain the center x, y and side length w, h of a prediction box including a target object. Finally, the output layer outputs these parameters with its confidence.

4. Experiment

We prepared 2106 images for testing. There are white-cane users in the 1053 images, while not in the other 1053 images. The threshold for the confidence is set to be 0.4. The confusion matrix is shown in Table 1, and the detection accuracy is shown in Table 2.

Table 1. The confusion matrix

	White cane detected	White cane not detected
Image with white-cane	781	272
Image without white-cane	10	1043

Table 2. Detection accuracy

Precision	0.987
Recall	0.734
F-score	0.842

5. Conclusions

This research proposed a detection method of white-cane users based on surveillance cameras and YOLO. The proposed method was applied to 1053 and 1053 images with and without white-cane users, respectively. The detection accuracy was evaluated with the F score, which was 0.842.

The main task in the future is to reduce the false detection rate for objects that are very similar in shapes to white-canes, such as umbrellas and normal sticks.

6. ACKNOWLEDGMENT

This work was supported in part by the JSPS KAKENHI Grant Number 19H04500.

References

- (1) Ministry of Land, Infrastructure, Transport and Tourism, Survey on Difficulties in Daily Life (National Survey on Children and Persons with Disabilities at Home), 2018
- (2) Koji Nakamura, NHK Commentary Archives: "Successive Station Home Fall Accidents: Protect the Visually Impaired" (Public Theory), 2020
- (3) Ministry of Land, Infrastructure, Transport and Tourism, Status of Installation of Platform Doors (March 31, 2020), 2020
- (4) J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- (5) Tzutalin. Labellmg. Git code (2015). <https://github.com/tzutalin/labellmg>
- (6) Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe and Youngjoon Yoo. "CutMix: Regularization Strategy to Train Strong Classifier with Localizable Features." arXiv preprint arXiv:1905.04899(2019).
- (7) Hongyi Zhang and Moustapha Cisse and Yann N. Dauphin and David Lopez-Paz," mixup: Beyond Empirical Risk Minimization", arXiv:1710.09412(2017)
- (8) ultralytics. YOLOv5. Git code (2021) <https://github.com/ultralytics/yolov5>
- (9) Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun," Deep Residual Learning for Image Recognition", arXiv:1512.03385(2015)
- (10) Alexey Bochkovskiy and Chien-Yao Wang and Hong-Yuan Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection", arXiv:2004.10934(2020)
- (11) Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8759-8768