Disabled Sign Recognition Using Single Shot Detection FPN

Rakhmonov Akhrorjon Akhmadjon Ugli*, Barathi Subramanian*,
Taehun Kim**, Jeonghong Kim*
*Kyungpook National Univ., **Dipvision

r.akhror@knu.ac.kr, achu_samriti@yahoo.com, dipvision.ceo@gmail.com,
*jhk@knu.ac.kr

Abstract

We will propose a model that can effectively detect and recognize disabled signs on the windshield of cars. Since this problem needs solution primarily in real time, utilizing mobile and embedded devices Mobilenet along with Feature Pyramid Network(FPN) will be leveraged. Therefore, the usage of depthwise separable convolution(DSC) and feature extractor with pyramid concept of FPN along with Single Shot Detection(SSD) will contribute to the accuracy while maintaining high speed of proposed model. For verification the effectiveness of our approach, a dataset with disabled sign on the car windshields will be collected and annotated, then conducted experiments will show us better results in the validation data compared to the model without FPN.

Ⅰ. Introduction

In general, handicap parking spaces in parking lots provide the most convenient and easy to access the entrance of the building. For humans it is a time-consuming task to identify the disabled sign on the windshield of vehicles. Therefore, we propose an algorithm that mitigates these issues by introducing the idea of using DSC for the reduction of mathematical operations, SSD for faster speed and FPN to improve the accuracy of the predictions.

The limited dataset means that we should exploit transfer learning in order to prevent overfitting and decrease model training time. We decided to leverage FPN since SSD only performs much worse for small object detections. However, FPN is not an object detector model despite being very effective tool at detecting small objects, it is conjugated to object detection models. SSD is more suitable for our task than Regional Proposal Networks (RPN) based approaches that must take two shots. Also, B.Subramanian et al.[1] proved the effectiveness of SSD over two shot approaches but in a different application from ours.

The sequence of the paper is organized as follows. In Section 2 we discuss some other available methods for the same task as ours, while Section 3 explains our proposed methodology. And we show our experimental results in Section 4. Then we represent conclusions along with future work in Section 5.

Ⅱ. Related Work

Many studies have made a comparison of object detection algorithms that specifically detect one type of object. S.Choyal and A.K.Singh[2] carried out the experiment on the performance of Faster-RCNN on traffic sign detection. According to the results, Faster-RCNN showed high accuracy but requires long training time since it possesses a lot of parameters to train.

Similarly, research conducted by C.Kumar B. et al.[3] concentrated on multiple object detection on Closed-Circuit Television(CCTV) cameras using YOLO object detection algorithms. Since their work is lack of explanation on their experimental method and they used unstandardized metrics, the comparison results are considered unsatisfactory.

As a matter of fact, we get inspiration from the work of T.Lin et al.[4] introducing FPNs to show significantly improved results when compared to other State-of-the-Art methods but again with Faster-RCNN which we already mentioned that it requires longer training time than ours.

Taking into consideration disadvantages of the algorithms above, we propose fast and reasonably accurate real-time disabled sign detection algorithm leveraging transfer learning.

Ⅲ. The Proposed Methodology

Model Architecture: We designed a model that is illustrated in Figure 1, based on SSD Mobilenet which shows accurate result and efficient effect. In addition to that, we apply FPN to improve model performance while detecting small size objects.

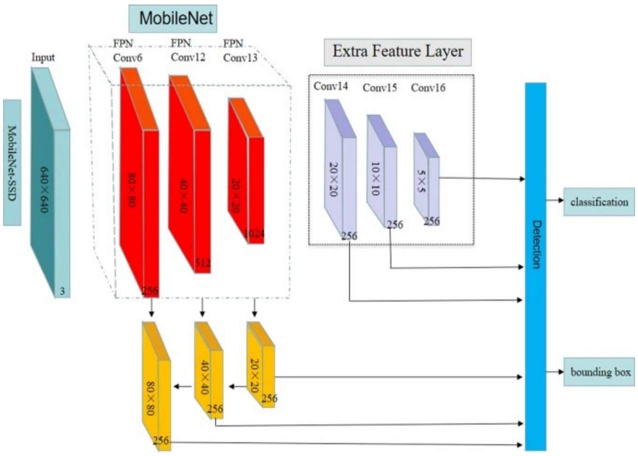
Method: Our proposed method consists of 4 stages:

1. Data collecting
2. Data Pre-Processing
3. Training the model
4. Evaluation of the model

Data collecting: In this paper we use dataset that we obtained with the help mobile devices. The dimension and quantity of images are 1920 x 1080 and 1025 respectively. We split the dataset into two sets: train and test data, which can be represented as 90% and 10% of the whole data accordingly.

Data Pre-Processing: In this stage, first we perform the process of cropping so that the focus on the sign is preserved. To crop the disabled sign and other signs from the image, we have used a graphical annotation tool called labelImg and draw a bounding box around them.

Training the model: Our initial model was pre-trained on coco dataset which is a large-scale image dataset containing 328,000 images of various objects as well as humans. We chose the batch size of 4 considering limited dataset, activation function is RELU[5], optimizer we use is momentum optimizer and number of steps is 10,000. Pre-processed image is inputted into the Mobilenet model leveraging FPN.



(Figure 1) The architecture of our model

Evaluation of the model: In the final stage we implement evaluation of our model with test data. For this process we take 10% of the overall dataset and check graphically illustrated outputs on which bounding boxes with respective recognition text is shown.

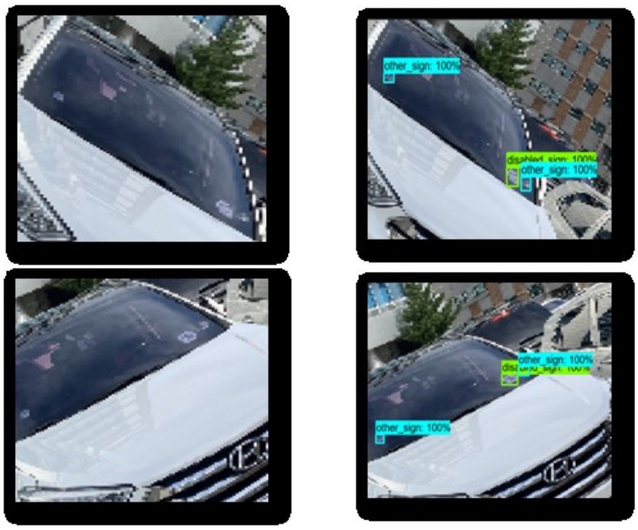
IV. Expiremental Results

For implementing the proposed model, we used Python 3.9.13 version on a personal computer with 32GB RAM and Intel i5 2.90GHz CPU, Windows 10 of 64-bit operating system. For evaluation of the performance of our model, we used Loss, which includes Localization and Classification loss, mean average precision(mAP) and average precision(AR) metrics as shown in Table 1. In just 10,000 steps our model for identifying the disabled sign outperformed similar model without FPN in all cases of given metrics which gives us confidence that with more steps we can assume it will show even more convincing performance.

(Table 1) Performance comparison

Network Model	Loss	mAP	AR
SSD Mobilenet	3.784	0.11	0.01
SSD Mobilenet FPN	0.595	0.21	0.03

Random input images and their corresponding output images with signs detected and recognized by our model are illustrated in Figure 2.



(a) (b)

(Figure 2) (a)input images (b)output images

V. Conclusion and Future Work

In this paper, we proposed a SSD model with applying FPN technique for disabled sign recognition and detection. Our experimental results revealed that in only 10,000 steps our model was able to outperform the one without FPN concept. We will continue to expand our dataset and further improvement on the SSD algorithm can be accomplished to get even more speed and accuracy.

ACKNOWLEDGMENT

This work was supported by project for Collabo R&D between Industry, University, and Research Institute funded by Korea Ministry of SMEs and Startups in 2022.(S3245927)

참 고 문 헌

[1]Subramanian B., Olimov B., Kim J.H., " Real-time hand gesture recognition for Indian sign language," 2021.

[2]Choyal S., Singh A.K., "A review of research on object detection based on deep learning," J. Phys. Conf. Ser., vol. 1684, 2020.

[3]Kumar, B., Punitha R. and Mohana., "YOLOv3 and YOLOv4: Multiple object detection for surveillance applications," 3rd ICSSIT, pp. 1316-1321, Aug. 2020.

[4]Lin T., Girshick R., He K., Harihan B. and Belongie S., "Feature pyramid networks for object detection," Intl. Conf. Parallel Distrib. Process, pp. 1500-1504, 2019.

[5]Agarap A.F., "Deep Learning using Rectified Linear Units (RELU)," March 20156