A Traffic Sign Detection Algorithm Based on Deep Convolutional Neural Network

Xiong Changzhen, Wang Cong, Ma Weixin, Shan Yanmei
Beijing Key Laboratory of Urban Intelligent Traffic Control Technology
North China University of Technology
Beijing, China
e-mail: xczkiong@163.com, 357597466@qq.com

Abstract—**Traffic sign detection plays an important role in driving assistance systems and traffic safety. But the existing detection methods are usually limited to a predefined set of traffic signs. Therefore we propose a traffic sign detection algorithm based on deep Convolutional Neural Network (CNN) using Region Proposal Network(RPN) to detect all Chinese traffic sign. Firstly, a Chinese traffic sign dataset is obtained by collecting seven main categories of traffic signs and their subclasses. Then a traffic sign detection CNN model is trained and evaluated by fine-tuning technology using the collected dataset. Finally, the model is tested by 33 video sequences with the size of 640×480. The result shows that the proposed method has towards real-time detection speed and above 99% detection precision. The trained model can be used to capture the traffic sign from videos by on-board camera or driving recorder and construct a complete traffic sign dataset.**

Keywords-traffic sign; object detection; convolutional neural network; region proposal network; dataset

I. INTRODUCTION

Traffic sign is significant security facilities on the road, which plays an important role in regulating traffic behavior, ensuring the safety of the road and guiding the smooth passage of vehicles and pedestrians, etc. As part of the intelligent transportation system, the detection of traffic sign is significant for driving assistance system, traffic sign maintenance, autonomous driving and other spaces. There have a number of research works done for traffic sign detection in the world. But most of the works are only for certain categories of traffic signs, for example, the speed limit sign[1], [2]. Traffic sign detection is generally regarded as challenging due to various complexities, for example, diversified backgrounds of traffic sign images. On the other hand, Chinese traffic signs are much more complex compared with western countries due to the large number of Chinese characters. How to efficiently detect all Chinese traffic sign has not been discussed. With the current understanding, there are not many research institutions that provide download of data and annotation. And the traffic signs are mainly confined to Europe [2]. Though many researches have been done for Chinese traffic sign detection and recognition, there is still not a common data set of Chinese traffic sign for researchers to download.

Therefore, in this paper, we propose a traffic sign detection algorithm based on deep convolutional neural network. The method can detect all 7 main categories of traffic signs in China. It takes only about 51.5 ms average detection time for one frame with the resolution of 640×480 video sequence. And it is robust to various interferences.

II. RELATED WORK

Recently, deep convolutional neural network have been a huge success in object detection. R. Girshick proposed the rich feature hierarchies for accurate object detection and semantic segmentation(RCNN)[3],which achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals. But it is slow because it performs a ConvNet forward pass for each object proposal, without sharing computation. Spatial pyramid pooling networks (SPPnets) [4] were proposed to speed up it by sharing convolutional feature map. The SPPnet method only computes a convolutional feature map for the entire input image. SPPnet also has notable drawbacks. Its training is a multi-stage pipeline and not end-to-end method. Fast R-CNN [5] enables end-to-end detector training on shared convolutional features and shows compelling accuracy and speed. The Fast R-CNN still uses the Selective Search method [6] to generate 2000 regions proposals. Faster R-CNN [7] introduces a training scheme that alternates between fine-tuning for the region proposal task and then fine-tuning for object detection. YOLO [8] reframes object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. SSD [9] using a small convolutional filter on feature maps to predict object categories and offsets in bounding box locations to improve the performance of object detection.

Inspired by these works, traffic signs as a specific categories of object detection can be expected to obtain a better performance than ever before. R. Qian et al. used region based convolutional neural network(R-CNN) for detection and recognition of China's prohibition, mandatory and warning 3 classes of traffic signs[10]. J. Jin et al. designed a hinge loss stochastic gradient descent optimization method to train a 3 layers network for traffic signs recognition [11]. The proposed methods usually limited to a predefined set of traffic signs because of the shortage of the traffic signs data.

At present, relatively well-known public traffic sign detection dataset mainly includes German traffic sign detection benchmark (GTSDB) dataset, Germany traffic sign recognition (GTSRB) dataset, and Belgium traffic sign (Belgium TS) dataset. The GTSDB dataset contains 900 high-Resolution natural scene images of the traffic signs, each image is 1360 × 800 pixels, the size of the traffic signs

978-1-5090-2377-6/16/$31.00 ©2016 IEEE

from 16 × 16 pixels to 128 × 128 pixels. It uses 600 images for training, and 300 images for testing. The GTSRB dataset contains 39209 images for training and 12630 images for testing. The signs are divided into three main categories including 43 subclasses Germany's traffic signs with sizes from 15 × 15 pixels to 250 × 250 pixels. The Belgium traffic signs dataset (BelgiumTS) is divided into two large datasets. One is detection dataset (BTSD) and the other is classification dataset (BTSC). The BTSD dataset contains 5905 train images and 3101 test images. The BTSC dataset contains 4591 train images and 2534 test images, the traffic signs in BelgiumTS are in 3 main categories including 63 subclasses. These traffic sign datum are mainly confined to Europe traffic sign. There is no relevant institutions have developed a large-scale, comprehensive public Chinese traffic sign dataset for researchers to download.

Therefore, we trained CNN model which can obtain images containing any of the traffic signs from videos captured by on-board camera or driving recorder. It can be used to construct a dataset fully contain all seven categories of Chinese traffic sign through simple post processing. It will have great value in the driving assistance systems and traffic safety applications.

III. TRAINING TRAFFIC SIGN MODEL

Faster R-CNN[6] introduces a novel region proposal Network(RPN), it shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals greatly improving the detection efficiency. Compared with several other latest object detection networks, region proposal network has certain advantage in cost of train and detection accuracy. Inspired by it, this paper propose a traffic sign detection algorithm based on deep convolutional neural network (CNN) using Region Proposal Network (RPN). It includes two main stages. One stage is to prepare train and evaluation dataset. The other stage is to fine-tuning pretrained CNN model.

A. Dataset

In order to achieve the detection of all 7 main categories of the Chinese traffic sign, we have collected traffic sign data from the internet and some photographs of street scenes in Beijing which containing Chinese 7 main categories and their subclasses. According to the related contents of Chinese national standard GB5768-2009 "The Road Traffic Signs and Marking" [12], Chinese traffic signs are divided into 7 main classes. They are warning sign, prohibitory sign, mandatory sign, guide sign, tourism sign, road construction safety sign and auxiliary sign. And each category is subdivided into a number of different subclasses. The details are shown in table I.

The data used for training traffic sign model consists of pictures from the internet and some photographs of street scenes in Beijing. The initial dataset contains 726 images and 1847 traffic signs, including Chinese 7 main categories of traffic sign and their subclasses. Calibration tool is used to generate the groundtruth of the traffic signs for each image. In order to increase the amount of training data and improve generalized ability of the model, we processed our initial dataset by motion blur and 12 different levels of brightness changes. Finally a dataset with 726×26=18876 images contain 48022 traffic sign data is obtained. Then we randomly select three-quarters of the data(14170 images) for training and validating, the rest of the quarter(4706 images) for testing.

TABLE I. THE CLASSES OF THE TRAFFIC SIGNS IN CHINA

categories	subcategories	characteristic
warning sign	42 classes	regular triangle, yellow bottom, black border and black pattern
prohibitory sign	42 classes	circle, white bottom, red side, and black pattern
mandatory sign	29 classes	circle or rectangle, blue bottom and white pattern
guide sign	62 classes	Rectangle, blue or green bottom and geographic information
tourism sign	17 classes	Rectangle, brown bottom, white pattern and sight spot's information
road construction safety sign	26 classes	Rectangle, blue bottom, white border, yellow diamond pattern, text
auxiliary sign	--	Rectangle, white bottom, black text or pattern, black border

B. Training and Testing Traffic sign model

We use the Faster R-CNN method to fine-tune the traffic sign detection model by end-to-end training method. The pretrained model on ImageNet is applied to to initialize the network weights. The ROI data layer and cls_score layer's output is set as 2(sign or not sign), and the bbox_pred layer's output is set as 8(2×4).The parameters of the train network is shown in Table II.

TABLE II. PARAMETER CONFIGURATION

parameter	Value
base_lr	0.001
lr_policy	Step
gamma	0.1
stepsize	50000
display	20
average_loss	100
momentum	0.9
weight_decay	0.0005
iter_size	2

We trained three traffic sign models VGG16, VGG_CNN_M_1024 and ZF. Then the three model is tested by the testing dataset(4706 images). The Mean Average Precision(mAP) and the average test time in different number of iterations for three trained model are shown in Table III.

TABLE III. COMPARISON OF DIFFERENT NETWORK

Models	Iterations	Test Data	MAP	Average Test Time
VGG16/VGG_CNN_M_1024/ZF	50000	4706	0.9086/0.9068/0.9077	160/71/60
VGG16/VGG_CNN_M_1024/ZF	60000	4706	0.9088/0.9077/0.9080	160/71/60
VGG16/VGG_CNN_M_1024/ZF	70000	4706	0.9089/0.9081/0.9083	160/71/60
VGG16/VGG_CNN_M_1024/ZF	80000	4706	0.9088/0.9077/0.9085	160/71/60
VGG16/VGG_CNN_M_1024/ZF	90000	4706	0.9089/0.9078/0.9083	160/71/60
VGG16/VGG_CNN_M_1024/ZF	100000	4706	0.9088/0.9079/0.9087	160/71/60

From Table III, we know that the traffic sign detection model's accuracy tends to be stable after 70000 iterations. The three models have the closed mAP 0.91. The ZF model has the highest detection efficiency, its average detection time is about 60ms in NVIDIA GTX980Ti GPU(6GB memory) hardware environment.

IV. TRAFFIC SIGN DETECTION IN VIDEOS

In order to verify our trained model, 33 video sequences captured by on-board camera and mobile phone are used to test them. The sequences include ordinary city streets and highways, and containing day time and night time different illumination conditions, and any other Interference factors such as slight rotation, occlusion, Motion blur and so on. The resolution is 640×480. There are 1057 traffic signs in these videos. Different model's detection results are shown in Table IV.

TABLE IV. STATISTICS OF TRAFFIC SIGN DETECTION IN VIDEO

Models	Number of signs in videos	Number of detection	Average time per frame	Detection rate
VGG16	1057	1053	145 ms	99.62%
VGG_CNN_M_1024	1057	1049	60.5 ms	99.24%
ZF	1057	1050	51.5 ms	99.33%

The experimental results show that in the continuous image sequence the traffic sign detection rate is above 99%, and the ZF's average detect time is 51.5ms per frame, it is close to the real time detection speed.

The experimental results of various interferences including different illumination conditions, rotation, occlusion, motion blur and so on are shown in Fig. 1. The Fig. 1 shows that our method can effectively detect the Chinese all kinds of traffic signs and it is robust to various interferences.

(a). Different illumination detection results.

(b). Multiple signs detection results.

(c). Sign occlusion detection results.

(d). Twist sign detection results.0

Figure 1. Detection results of various interferences

V. CONCLUSION

This paper has presented a Chinese traffic sign detection algorithm based on deep convolutional neural network using region proposal network in Faster R-CNN. The method can detect all seven main categories Chinese traffic sign. We trained and compared three models. The experimental results show that the traffic sign detection rate of our algorithm is above 99% and the detection time is towards real-time in video sequences.

The trained model can be used to capture the traffic sign in the videos or images from the web and generate a complete dataset about traffic signs through simple post-processing, therefore this method is of great significance in terms of data collection of traffic signs.

In future, we will focus on how to obtain accurate groundtruth and the class of the sign detected automatically.

ACKNOWLEDGEMENT

This paper is supported by the Beijing Higher Education Young Talents Cultivation Project (CIT&TCD 201404009) and the Subject Construction Project of North China University of Technology (XN075).

REFERENCES

[1] Biswas, R., H. Fleyeh, and M. Mostakim, "Detection and classification of speed limit traffic signs," Proc. Computer Applications and Information Systems(WCCAIS), Jan. 2014, pp. 1-6, doi:10.1109/WCCAIS.2014.6916605.

[2] H. Liu, J. Li, X. Hu and F. Sun, "Recent progress in detection and recognition of the traffic signs in dynamic scenes," Journal of Image and Graphics. vol. 18, May. 2013, pp. 493-503.(In Chinese)

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June. 2014, pp. 580-587, doi:10.1109/CVPR.2014.81.

[4] K. He, X. Zhang, S. Ren and J. Sun , "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, Jan. 2015, pp. 1904-1916, doi: 10.1109/TPAMI.2015.2389824

[5] R. Girshick, "Fast r-cnn," Proc. IEEE International Conference on Computer Vision (ICCV), Dec. 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169

[6] J. R. Uijlings, K. E., de Sande, T. Gevers, A. W. Smeulders, "Selective search for object recognition," International Journal of Computer Vision (IJCV), vol. 104, March. 2013, pp.154–171, doi: 10.1007/s11263-013-0620-5.

[7] S. Ren, K. He,R. Girshick,J. Sun,"Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," arXiv:1506.01497, 2015, in press.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: unified, real-time object detection," arXiv:1506.02640, 2015, in press.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, "SSD: single shot multibox detector," arXiv:1512.02325, 2015, in press.

[10] R. Qian, B. Zhang, Y. Yue and Z. Wang, "Robust Chinese traffic sign detection and recognition with deep convolutional neural network," Proc. IEEE International Conference on Natural Computation(ICNC), Aug. 2015, pp. 791-796, doi: 10.1109/ICNC.2015.7378092.

[11] J. Jin, K. Fu and C. Zhang ,"Traffic sign recognition with hinge loss trained convolutional neural networks," IEEE Transactions on Intelligent Transportation Systems. vol. 15, Dec. 2014, pp. 1991-2000, doi:10.1109/TITS.2014.2308281.

[12] GB_5768-2009,"The Road Traffic Signs and Marking,"(In Chinese, Chinese national standard).