2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)

Road Surface Traffic Sign Detection with Hybrid Region Proposal and Fast R-CNN

Rongqiang Qian*, Qianyu Liu*, Yong Yue*, Frans Coenen† and Bailing Zhang*
*Department of Computer Science and Software Engineering
Xi'an Jiaotong-Liverpool University
Suzhou, P.R.China, 215123
†Department of Computer Science
University of Liverpool
Liverpool, L69 3BX, United Kingdom

***Abstract*—Detection of traffic signs plays an important role in autonomous driving, traffic surveillance and traffic safety. Previous research in Traffic Sign Detection (TSD) generally focused on traffic signs which are over the roads, the traffic signs on road surface have not been discussed. In this paper, we propose a road surface traffic sign detection system by applying convolutional neural network (CNN). The proposed system consists of two main stages: 1) a hybrid region proposal method to hypothesize the traffic sign locations by taking into account complementary information of color and edge; 2) feature extraction, classification, bounding box regression and non-maximum suppression by Fast R-CNN. Extensive experiments have been conducted using our field-captured dataset, demonstrating outstanding performance with regard to high recall and precision rate. The overall average precision (AP) is about 85.58%.**

***Keywords*:—Advanced Driver Assistance, traffic sign detection, deep learning, convolutional neural networks, Fast R-CNN.**

I. Introduction

The extraction of information from various traffic signs is a key component in many applications, such as autonomous driving, traffic surveillance and traffic safety. In general, a traffic sign recognition (TSR) system involves two related issues: traffic sign detection (TSD) and traffic sign classification (TSC). To be more specific, the former one concentrate on the localization of the targets in the pictures while the later one pays attention to identifying the labels of the detected targets. Lots of outstanding results for the detection and classification of traffic signs have been proposed in [1], [2], [3], [4], [5], [6], [7].

Purposing to be easily noticeable by drivers and pedestrians, traffic signs are usually designed to have rigid and simple shapes as well as uniform and attractive colors. Generally, traffic signs are erected at the side of roads, above roads or painted on the surface of roads. Most of TSR related works focus on traffic signs at the side of or above roads, and traffic signs on road surface have not been discussed. Different from traffic signs at the side of or above roads, the detection of traffic signs on road surface usually faces more challenges, because the problems such as scale variations, viewpoint variations, illumination conditions, motion-blur, occlusions and colors fading are generally much more severe.

Among the published works on traffic sign detection, the dominant approaches can be generally categorized into color-based method, shape-based method and machine learning-based method. The detection principle of color-based method and shape-based method rely on the two most distinguished attributes of traffic signs, namely, color and shape. Inspired by the successes of machine learning for face detection and pedestrian detection, AdaBoost and Support Vector Machine (SVM) have also been utilized in traffic sign detection.

Recently, the development of deep learning has attracted lots of attention in computer vision and pattern recognition research as more and more promising results are published on a range of different tasks. Benefit from the huge success achieved by R-CNN [8], object detection based on CNN is significantly developed. Moreover, by introducing SPPnet and multi-task loss, Fast R-CNN [9] achieves a much higher precision, meanwhile, the training and testing speeds have been boosted up to 10 times faster.

The main motivation of this paper is to present a road surface traffic sign detection system. The main characteristics of our proposed system include:

- A hybrid region proposal method which takes into account the complementary information of color and edge.
- Fast R-CNN to perform classification and bounding box regression simultaneously.

The rest of this paper is organized as follows: Section 2 outlines some related research on traffic sign detection and CNN applications; Section 3 gives a detailed introduction of the proposed TSD system; Experimental results will be provided in Section 4, followed by conclusion in Section 5.

II. Related Works

A. Traffic Sign Detection

TSR has been researched for many years. Depending on which priori knowledge is adopted, color-based methods and geometry-based methods are the most widely used approaches. Color-based methods have low computational costs and strong robustness to projective distortion [1], [3], [2]. As geometric shape is another important cue for detection of traffic signs, lots of works have proposed along this line, including Radial Symmetry Detector [4] and Triangular Detector [5]. TSD has

978-1-5090-4093-3/16/$31.00 ©2016 IEEE

also become inseparable with machine learning. For example, a real-time detection scheme was proposed in [6], and an aggregate channel features (ACF) and integral channel features (ICF) detector was reported in [7]. However, a fundamental shift has occurred in the past years, with sliding-window being replaced by region proposals. The main motivation is to generate candidate object proposals which would capture most of the objects in an image. Some influential region proposal algorithms include [10], [11], [12]. Follow this line, Greenhalgh et al. [3] employed MSERs for region proposal. A hybrid region proposal method which consists of MSERs and Wave-based Detector (WaDe) was reported in [2].

B. Convolutional Neural Network

Due to the power of representational learning from raw data, deep learning [13], [14] has acquired general interests in recent years. Among the deep learning models, recent popularity of CNN has dominated various computer vision problems, and object detection and recognition in particular. Krizhevsky et al. [15] proposed a large-scale deep convolutional network trained by standard back propagation, with a milestone on the large-scale ImageNet object recognition dataset [16], attaining a significant gap compared with existing methods that adopt shallow models.

Our work on road surface traffic signs detection was mainly inspired by one of the most influential CNN models, namely called Fast R-CNN, proposed by Girshick et al. [9] for object detection. By selective search [10], a mass of candidate bounding boxes will be produced as the proposal regions, which are subsequently labeled and regressed by Fast R-CNN [9] simultaneously. Different from Fast R-CNN, selective search [10] is not adopted in our system, because the uniform and distinctive colors of traffic signs can be exploited to simplify the generation of proposal regions.

III. Road Surface Traffic Sign Detection

The whole detection system overview is demonstrated in Fig. 1. In the first stage of our system, a large number of candidate regions that may contain the desired objects are generated by region generators. In the second stage, all the generated regions will be passed to Fast R-CNN [9] with following processing steps: 1) feature extraction with CNN; 2) classification and bounding box regression by multi-task neural network; 3) non-maximum suppression for the final outputs.

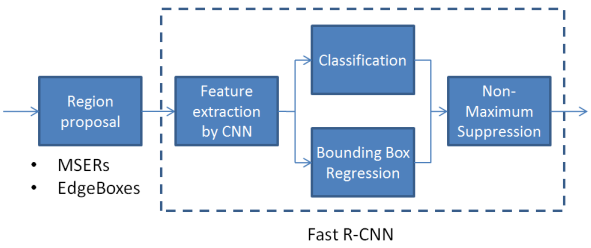

Fig. 1. System overview

A. Region Proposal

Instead of using sliding-window, our end-to-end road surface traffic sign detection system generates input boxes by region generators. Therefore, a fast and high recall is pursued in the initial region proposal stage while high precision will be achieved in later stages. In order to evaluate the recall rate for a region proposal method, a bounding box is deemed to be true if its overlap with ground-truth bounding box is sufficiently high. The overlap for two bounding boxes, denoted as b_1 and b_2, can be defined as the ratio of intersection over union (IoU): $\frac{b_1 \bigcap b_2}{b_1 \bigcup b_2}$ [17].

The detection targets include a number of different traffic signs, as illustrated in Fig. 2. Due to the problems such as scale variations, viewpoint variations, illumination changes, motion-blur, occlusions and colors fading, targets may have huge difference in shape, scale and color. Therefore, in our system, we combine bounding boxes from two proposal algorithms, namely, MSERs [11] and EdgeBoxes [12].

Fig. 2. Illustration of road surface traffic signs

1) Maximally Stable Extremal Regions: MSERs [11] denote a set of outstanding regions that are proposed in an input image. The extremal property of these regions are defined by its intensity function. Moreover, MSERs are scale-invariant and affine-invariant. Due to these advantages, MSERs have been broadly applied in various detection tasks, usually with regions generated by binarization and connected components analysis. More specifically, an input image is firstly binarized using different threshold levels, and then connected components analysis is used to find all the connected components at each threshold level. Finally, the regions that maintain their shape and area at several thresholds are selected as MSERs.

2) EdgeBoxes: Recently, EdgeBoxes [12] has shown prominent performance comparing with other methods. The key intuition of EdgeBoxes is that: the number of contours wholly covered by a bounding box is a cue of the possibility of containing object. To be more specific, objects usually

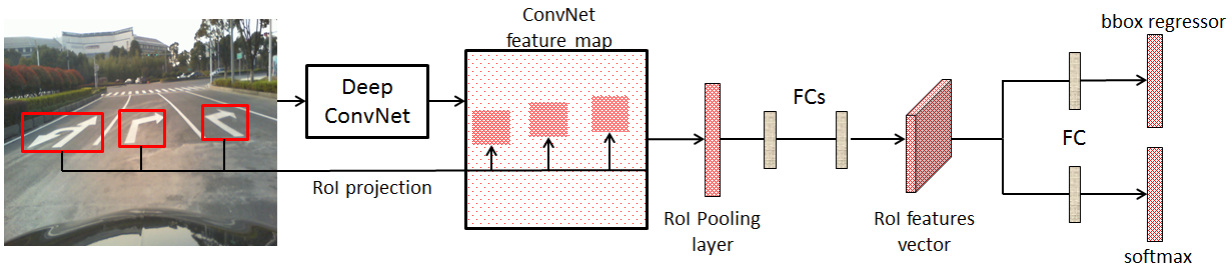

Fig. 3. System overview of Fast R-CNN

have obvious and enclosed edges. Following this intuition, a bounding box that contains enclosed edges tends to have object inside. As road surface traffic signs have distinguished and sharp boundaries, it seems to be particularly true that objects are composed of plentiful boundaries. Following the pipelines in [12], the Structured Edge Map detector is exploited to compute the edge response map, and non-maximum suppression is applied orthogonal to edge response in order to sparsify the edge map. To a candidate bounding box b, a score S_b is assigned, based on the number of edges wholly contained by in bounding box b.

B. Classification and Localization by Fast R-CNN

R-CNN based object detection approaches have demonstrated state-of-the-art performance for a variety of applications. However, R-CNN has its bottlenecks in a number of aspects: (i) high computational cost in training and testing stages; (ii) huge storage capacity request in training stage; (iii) separated training for classification and bounding box regression. In order to solve these problems, an improved scheme has been designed and proposed, which is called Fast R-CNN [9]. With the help of multi-task learning and regions of interest (RoIs) pooling, the training and testing speed is significantly improved, meanwhile, detection accuracy is also benefited from the new paradigm.

The system overview of Fast R-CNN is illustrated in Fig. 3. Different from R-CNN, Fast R-CNN takes whole images as input. Therefore, a feature map of entire input image is extracted at the final layer of CNN, with the addition of position of each region proposal, the feature of each regions is obtained. And then, all the feature maps are further processed by a RoI pooling layer, which is a special case of the spatial pyramid pooling layer, so that image size of region proposal can be arbitrary. Finally, Fast R-CNN performs classification and bounding box regression simultaneously by a multi-task network.

IV. Experiment

In this section, firstly, implementation details of our system will be introduced, and then the collected road surface traffic sign dataset will be presented. After this, the experiments results of proposed hybrid region proposal method will be illustrated and analyzed. Finally, the experiment results of whole system will be provided and discussed, followed by detection error analysis. The details will be introduced with the corresponding experiments in the following.

A. Implementation Details

To implement our TSD system, a computer with Xeon E3-1231 V3 CPU, 32GB memory and 6GB memory 970m GPU is employed. The program runs on a 64-bit Open-source Linux operating system with CUDA 7.5, Python 2.7.3, Matlab 2014b and Caffe deep learning platform installed. The proposed hybrid region proposal method is implemented with the published source code [12] in Matlab, and Fast R-CNN models are trained on the Caffe platform.

B. Field-captured Road Surface Traffic Sign Dataset

The dataset was recorded by a camera set up in a vehicle with a fixed shooting angle. The captured videos have a resolution of 1280*720 pixels. We decompressed the videos and collected all the frames that contain traffic signs. The total number of the collected images is 2223 with 4204 traffic signs included, with details shown in Table I. Some of the examples are illustrated in Fig. 4. The sizes of traffic sign in the images vary from 28×229 to 790×410. The ground-truth boxes were subsequently labeled. Then, candidate regions were generated by applying MSERs and EdgeBoxes. Finally, training and testing data were collected.

TABLE I
DETAILS OF FIELD-CAPTURED DATASET

	Image	Traffic signs
Training dataset	1606	2996
Testing dataset	617	1208
Total	2223	4204

C. Performance Evaluation of Region Proposal

The recall rates for region proposal stage are shown in Fig. 5. It is obvious that the combination of the two proposal methods shows significantly boosted performance. When the value of IoU is set to 0.5, the recall rate for EdgeBoxes, MSERs and the combination of the two methods are 48.34%, 87.83% and 89.74% respectively. The missed proposal regions

Fig. 4. Samples from our field-captured dataset

are mainly caused by scale variations, changes of illumination conditions and colors fading.

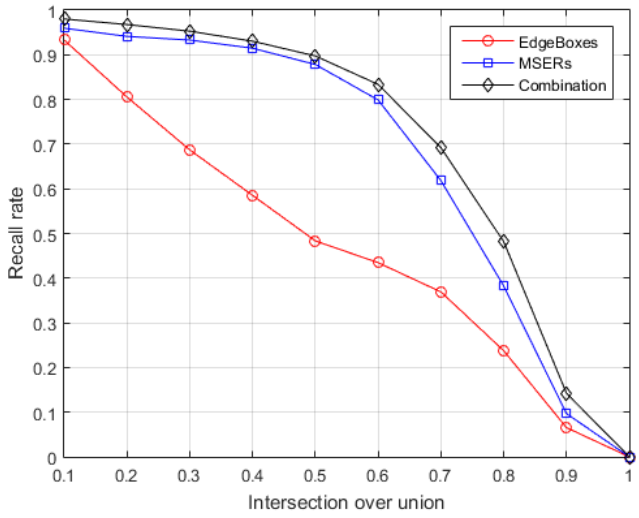

Fig. 5. Illustration of the recall rates with the changes of intersection over union (IoU) between the range (0.1,1).

D. Performance Evaluation of Detection System

In order to provide a comprehensive evaluation, two experiments were set up for testing the performance of multi-task training and bounding box regression. The precision-recall (PR) curves for whole proposed system are demonstrated in Fig. 6. As the figure indicates, the highest recall rates achieved by Fast R-CNN with or without bounding box regression are 90.73% and 88.33% respectively, meanwhile, the corresponding precision rates are 14.49% and 71.23% respectively. Comparing with the recall rate 89.74% obtained in region proposal stage, Fast R-CNN with bounding box regression attains 0.99% improvement and Fast R-CNN without bounding box regression loses 1.41% positive regions, this result indicates that bounding box regression indeed improves performance, regions with low IoU value: $(0.1 - 0.5)$ can be regressed to have higher IoU value: (> 0.5). On the other hand, by introducing multi-task training for bounding box regression, the precision rate is smaller than single-task CNN, therefore, multi-task training has negative influence to classification performance.

The average precisions (APs) of Fast R-CNN with or without bounding box regression have been shown in Table II, which are 85.58% and 83.99% respectively. So that multi-task training in Fast R-CNN enhances the performance of whole system.

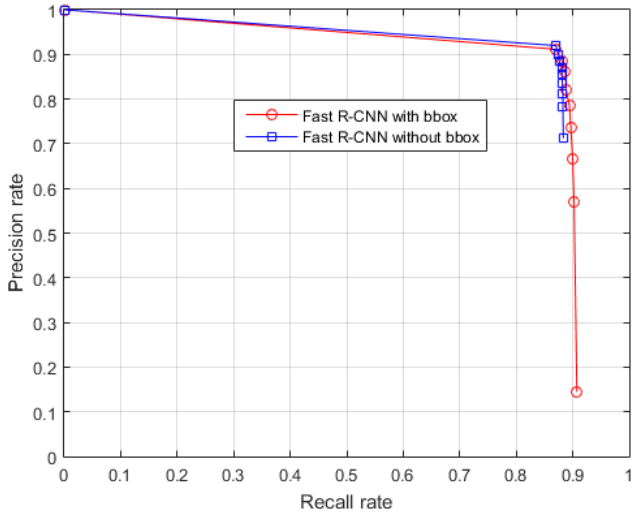

Fig. 6. Illustration of precision vs. recall curve for the whole proposed system.

TABLE II
DETECTION RESULTS

Method	AP
Fast-RCNN with bbox	85.58%
Fast-RCNN without bbox	83.99%

E. Error Analysis

Although high performance has been achieved on our road surface traffic sign dataset, there still exists missing detections, which are mainly caused by scale variations, changes of illumination conditions and colors fading. Some of the examples are shown in Fig. 7. Since the recall rate of regions proposal is only about 89.74%, region proposal is the bottleneck of our system. This means that more accurate region proposal algorithms will be focused to further improve the overall detection performance.

Fig. 7. Illustration of examples where the detection algorithm failed.

V. CONCLUSION

In this paper, a road surface traffic sign detection system is proposed, with main contributions including: (i) A hybrid region proposal method which takes into account the complementary information of color and edge; (ii) Fast R-CNN to perform classification and bounding box regression simultaneously. By combining two popular region proposal methods, namely, MSERs and EdgeBoxes, recall rate is significantly improved. As the features extracted from CNN are more discriminative, high recall and precision rates of the entire system have been achieved. Extensive experiments have been performed using our dataset, yielding promising results.

REFERENCES

[1] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, and S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 4, pp. 917–930, Dec 2010.

[2] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, and L. D. Stefano, "Traffic sign detection via interest region extraction," *Pattern Recognition*, vol. 48, no. 4, pp. 1039 – 1049, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320314002179

[3] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 1498–1506, Dec 2012.

[4] N. Barnes, A. Zelinsky, and L. Fletcher, "Real-time speed sign detection using the radial symmetry detector," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 322–332, June 2008.

[5] R. Belaroussi and J.-P. Tarel, "Angle vertex and bisector geometric model for triangular road sign detection," in *Applications of Computer Vision (WACV), 2009 Workshop on*, Dec 2009, pp. 1–7.

[6] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, June 2005, pp. 255–260.

[7] A. Mogelmose, D. Liu, and M. Trivedi, "Detection of u.s. traffic signs," *Intelligent Transportation Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–10, 2015.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, June 2014, pp. 580–587.

[9] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1440–1448.

[10] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013. [Online]. Available: https://ivi.fnwi.uva.nl/isis/publications/2013/UijlingsIJCV2013

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004, british Machine Vision Computing 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885604000435

[12] C. Zitnick and P. Dollar, "Edge boxes: Locating object proposals from edges," in *Computer Vision ECCV 2014*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8693, pp. 391–405. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10602-1-26

[13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 5786, p. 504, 2006.

[14] G. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, July 2006.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 248–255.

[17] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010. [Online]. Available: http://dx.doi.org/10.1007/s11263-009-0275-4