2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Traffic-Sign Spotting in the Wild via Deep Features

Jinkang Guo, Jianyun Lu, Yanyun Qu*, Cuihua Li
Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Information Science and Engineering, Xiamen University, 361005, China
{jinkang.guo, sikyadjy, quyanyun}@gmail.com, chli@xmu.edu.cn

***Abstract*—This paper focuses on traffic sign spotting (TSS) which automatically recognizes not only the conventional traffic signs but also information, facility and service signs, and traffic lights. TSS is divided into two sequential tasks: detecting traffic sign candidate regions in an image and recognizing the traffic signs in the regions. It is a very challenging task. We make the following contributions: 1) we create a traffic sign collection from the driverless car. The traffic signs are shot under the natural environment which covers large variation in illuminance and weather conditions. It not only contains the common traffic signs but also contains the information, facility and service signs which are called signposts, as well as traffic lights. 2) we proposed a systematic solution. We construct an Inception convolutional neural network. We use Faster–RCNN for traffic sign detection and make it suitable to detect small targets. 3) We adopt three schemes for the common traffic signs, the signposts and the traffic lights, respectively. The experimental results demonstrate the effectiveness and efficiency of our methods. Our methods won the first place in the traffic sign recognition task of Intelligent Vehicle Future Challenge 2017, China.**

Keywords—traffic sign; detection; recognition; Inception convolutional neural networks

I. INTRODUCTION

While traffic sign recognition is well studied and there are many available systems, the automatic detection and recognition of traffic sign within images are far less developed. However, traffic signs contained within images can be of great prompt information, and so is an important step towards both driverless car and autonomous navigation. Traffic sign spotting in the wild is usually divided into two tasks: traffic sign detection and traffic sign recognition. Traffic sign detection involves generating candidate bounding boxes that are likely to contain regions of traffic signs, while traffic sign recognition gets each candidate bounding box, and tries to recognize the traffic sign within it, or potentially reject the bounding box as a false positive detection.

Traffic sign spotting is a challenging task. The traffic signs are in a variety of appearances, with high inter-class similarity, and complicated background. Meanwhile, they accompany with occlusion, a variant of viewpoint, illumination, and so on. Moreover, traffic signs may only occupy a small fraction of an image, thus, they become small targets in the wild.

In this paper, we focus on a harder problem of spotting more extensive traffic signs than before. Fig. 1 shows some examples of traffic sign collection, which contains not only some common traffic signs such as warning signs but also the signposts and the traffic lights. This task is more difficult than only spotting the common traffic signs before, because the signposts are in a great variety of appearance and of shortage in the amount of current traffic sign datasets and the traffic lights in the wild are sensitive to the environment changes. Regarding the harder problem, we have created a traffic sign collection mainly from the competition video shot from the driverless car used in the Traffic Sign Recognition Competition 2015 and 2016. It provides more than 13000 images containing 77 traffic-sign instances. These images cover large variations in illuminance and weather conditions. Each traffic-sign in the collection is annotated with a class label, its bounding box. We call this benchmark Xiamen-Xi'an 13K. We show that a systematic method based on convolutional neural network can improve upon the performance of both the traffic sign detection and recognition tasks of this pipeline. To achieve this, we design an Inception Convolutional Neural Network. We adopt three schemes for the common traffic signs, the signposts and the traffic lights. In order to improve the detection accuracy of traffic signs, we train Faster-RCNN and tune the parameters in order to make it suitable for small targets. Our contributions are threefold:

1) We have created a new, more realistic traffic-sign collection. Compared to the widely used detection benchmark GTSDB which is a German traffic sign dataset used for the competition, our collection contains the signposts and the traffic lights, which are shot by a driverless car. The traffic signs cover the environment in the wild, with large variations in the aspects of illuminance and weather condition, as well as examples of occlusion.

2) We design a systematic method for an end-to-end traffic sign recognition. Regarding the large variation of traffic signs, we take three schemes to deal with the conventional traffic signs, the information, facility and service signs, and traffic lights, respectively.

3) We construct an Inception-CNN for multi-class traffic sign classification. And we use Faster-RCNN for traffic sign detection and make it suitable for small target.

In the following, we first introduce the related work in Section 2 and describe the implementation details of the proposed method in Section 3. Section 4 evaluates the proposed method on Xiamen-Xi'an 13K. Finally, conclusions are given in Section 5.

* Corresponding author: Yanyun Qu

Fig. 1. Traffic signs. Large variety, great similarity between classes.

II. RELATED WORK

Traffic sign spotting has become a popular problem in intelligent vehicles, and various methods have been proposed to address this challenging task. These methods are divided into two classes: the classical methods and deep learning based methods.

There are many classical end-to-end traffic sign recognition methods [1, 2, 3, 4]. In these classical methods, selective search is used to get proposal regions, and implemented by combining hand-crafted features and machine learning algorithm. The most common of hand-crafted features is SIFT [5, 6], HOG [7], LBP [8], etc. After describing traffic signs, classifiers are learned based on the feature representing for traffic sign detection and classification, such as support vector machines (SVM) [9], boosting [10], Bayesian classifiers [5] and random forest classifier [11].

Deep learning has achieved breakthrough recognition accuracy in a large number of applications in recent years, such as speech recognition, image classification and object detection. Considerable convolutional neural networks are constructed. AlexNet [12] firstly achieves success in image classification on ILSVRC2010. ResNet [13] improves the architecture of convolutional neural network by adding a shortcut skip and achieved better image classification accuracy. Inception convolutional neural network [14] changes the larger filter kernels to a combination of small filter kernels, which further improve the image classification accuracy. As for the deep learning based object detection, SPPNET [15] solve the problem that AlexNet requests the fixed input size. Fast R-CNN [16] improve the SPPNET by selecting preset pooling pyramid to reduce the proposal number of selective regions. Faster-RCNN [17] selects the proposal region in the feature map by using FPN and achieves good detection accuracy and speed. Though SSD [18] and YOLO [19] have better detection performance in general object detection, we experimentally find that Faster-RCNN is effective in small object detection.

Deep learning based traffic signs recognition is investigated recently. Wu et al. [20] transformed the original image into the gray scale image by using support vector machines, then use a CNN combined with fixed and learnable layers for detection and recognition traffic signs. Zhu et al. [21] trained an end-to-end CNN for simultaneously detecting and classifying traffic signs. Li et al. [22] proposed a Perceptual Generative Adversarial Network (Perceptual GAN) model to improve small object detection through narrowing representation difference of small objects from the large, and got a pretty good result in traffic sign detection. However, few of the existing methods discuss the recognition of signposts and sign lights and the common traffic sign recognition simultaneously.

In our work, we propose three schemes based on data characteristics to solve both traffic sign and traffic light detection and classification.

III. DATA COLLECTION

Unlike the German Traffic Sign Recognition Benchmark (GTSRB) [23], the Swedish Traffic Signs Dataset (STSD) [24] whose traffic-sign classes were limited, we add the classes of information, facility, service sign and the classes of traffic lights into the standard traffic sign classes. We build our traffic sign dataset on two traffic sign video sets: the 2015 Traffic Sign Recognition Competition Dataset (TSD-2015) and the 2016 Traffic Sign Recognition Competition Dataset (TSD-2016).

TSD-2015 is a 72-class traffic signal dataset used in the 2015 China Traffic Sign Detection and Recognition Competition. They are split into 7 main categories: (1) mandatory signs, (2) assist signs, (3) warning signs, (4) road construction safety signs, (5) tourism districts signs, (6) guided signs and (7) prohibitory or restrictive signs. There are great differences in appearance in traffic sign classes, especially in color, shape and scale. We build a dataset that contains 10611 training images and 8520 test images.

TSD-2016 is a 77-class traffic signal dataset used in the 2016 China Traffic Sign Detection and Recognition Competition. They are split into 5 main categories: (1) mandatory signs, (2) warning signs, (3) Traffic lights, (4) guided signs and (5) prohibitory or restrictive signs. There are great differences in appearance in traffic signal classes, especially in color, shape and scale. We built a dataset of 120 training videos and 50 test videos with 2965 training images and 936 test images.

We augment the training samples by randomly changing the brightness, transforming the contrast, and randomly adding the Gaussian noise. Data augmentation can effectively increase the diversity of training samples. Some examples of the data augmentation are shown in Fig. 2.

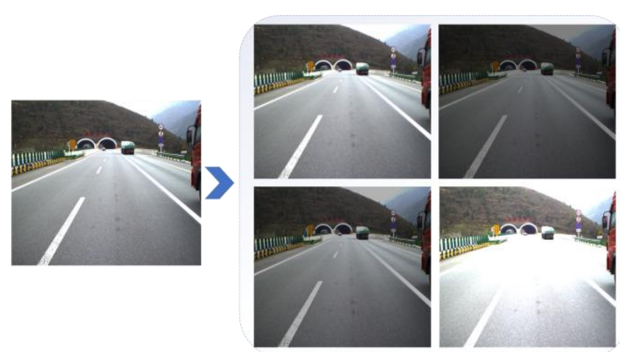

Fig. 2. Augmented samples

IV. THE PROPOSED METHOD

Similar to GTSRB, our traffic sign dataset contains three shapes: triangle, circle and rectangle, as shown in Fig. 3. Rectangle traffic signs include a subclass named signposts (Fig. 4) which are information, facility and service traffic signs. A signpost occupies a large region in an image, but the icon deciding its class label only occupies a very small region in the signpost region. As shown in Fig. 5, both signpost instances belong to the same class of 'Entrance', but there is a lot of interference information, so it is pretty difficult to subdivide the class. For traffic lights, in the wild, its color is greatly influenced by the natural environment such as illumination, weather, etc. As shown in Fig. 6, it is hard to tell which is the red light and which is the yellow light. It often makes classification confused.

In this paper, we proposed an end-to-end systematic method. Fig. 7 shows the framework of our method. In this framework, we divide the traffic sign spotting task into three tasks: traffic light spotting, signpost spotting and common traffic sign spotting, at the same time, we put forward three sub-frameworks based on Faster-RCNN and Inception-CNN to accomplish these tasks.

Fig. 3. Three types of traffic signs: triangle, circle and rectangle.

Fig. 4. Some examples of signpost signs

Fig. 5. Two different instances in the same class of 'Entrance'.

Fig. 6. Traffic lights in the wild.

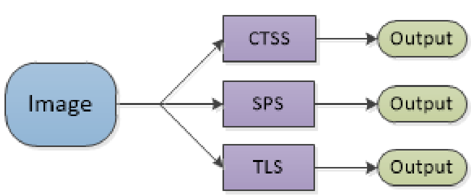

Fig. 7. An end-to-end systematic framework. CTSS: common traffic sign spotting framework. SPS: signpost spotting framework. TLS: traffic light spotting framework. Output: the classification probability of each class.

A. Faster-RCNN for traffic sign detection

Faster-RCNN includes the four basic steps of target detection: feature extraction, the candidate area generation, classification and location refinement into a deep network framework, greatly improving the running speed and its accuracy has reached the highest level. Although some other models can improve the speed of detection, such as YOLO, SSD, it is very hard to achieve the higher recognition accuracy than Faster-RCNN, especially in small target detection tasks. Therefore, according to the characteristics of the traffic sign detection task, it is a wise choice to implement Faster-RCNN on traffic sign detection.

As we said before, we divide the traffic sign spotting task into three tasks. And we train them respectively because of the difference in shape and size between the three kinds of data. We design the anchors [17] elaborately, the anchor ratio of the common traffic sign is set to [1:1], the anchor ratio of the signpost is set to [1:2, 2:1] and the anchor ratio of the traffic lights is set to [1:3, 3:1]. The anchor of common traffic signs and traffic lights is set to a small size and the anchor size of the signpost is set big. In this way, a strong detector can be trained to make full use of the shape and size of the traffic sign.

B. Classifier

In order to make classification tasks better, this paper proposes a classification model named Inception-CNN, as shown in Fig. 8, which uses a small filter kernel in the shallow layer and deepens the network. The deeper convolutional network makes it possible to obtain a large acceptable candidate region. Besides, the Inception structure makes the network wider which enables the multi-scale features to be combined optimally. In the network, we use 1 * 1 convolution kernel to reduce the feature maps' thickness and use two concatenation convolutional layers with convolution kernel size of 1 * 1 instead of the kernels of size 5 * 5 to reduce the number of parameters and speed up the computation.

We use a soft-max function as a classifier which is a multi-class logistic regression function. The soft-max function can compress a real number vector of k dimensions into a probability vector whose elements range from 0 to 1 and the sum of all the elements is 1. The probability of the target class is bigger, and the probability of non-target class is smaller. Let's denote X(i) as the probability of the ith class and $1\{X(i)\}$ as the indicator function. The probability of the jth class is formulated as,

$$P(X(j)|X) = \frac{exp(X(j))}{\sum_{i=1}^{k} exp(X(i))} \quad (1)$$

The cost function is the likelihood loss function,

$$J = -\sum_{i=1}^{k} 1\{X(i)\} \cdot logP(X(i)|X) \quad (2)$$

When a query image is tested, we use the index of the maximum probability in the probability vector as the target category.

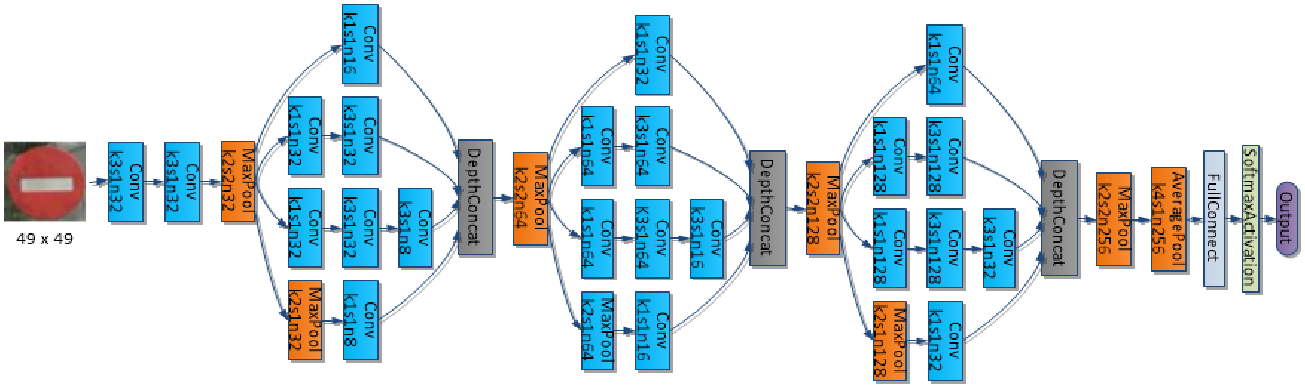

Fig. 8. Architecture of Inception-CNN with corresponding kernel size(k), stride(s) and number of feature maps(n).

C. Implementation details

1) Common traffic sign spotting(CTSS)

For common traffic signs, we propose a hierarchical traffic sign learning method which is shown in Fig.9. Firstly, we detect the traffic signs by Faster-RCNN according to three shapes: triangle, circle and rectangle. And then we implement Inception-CNN on each candidate region to distinguish the class label. Noted that there are some noise data in the three categories detected in shape, thus, a noise class is added.

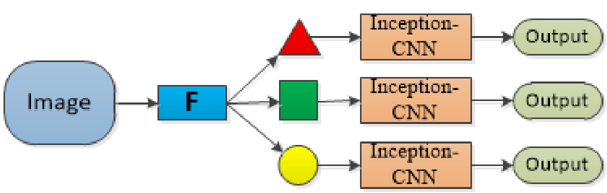

Fig. 9. Common traffic sign detection and classification framework. F: Faster RCNN framework. Output: the classification probability of each common traffic sign category.

2) Signpost spotting(SPS)

We detect the signposts by using Faster-RCNN. Since there is much interference information in the region of signposts, we do not treat the whole region of a signpost as the target, but only focus on the critical icon or text as the target, as shown in Fig. 10. We adopt a two-stage detection strategy which is shown in Fig. 11. Firstly, we detect the signpost as the rectangle shape by the Faster-RCNN in first stage, and then in the candidate region, we use the second stage Faster-RCNN to detect the critical icon or text. The main reason why we use two-stage detection rather than direct detection of key features is that some of the key features are similar to other traffic signs and are prone to confusion, as shown in Fig. 12.

Fig. 10. Key features of the signpost targets.

Fig. 11. Signpost detection and classification framework. F1, F2: Faster-RCNN framework at each stage. Output: the classification probability of each signpost class.

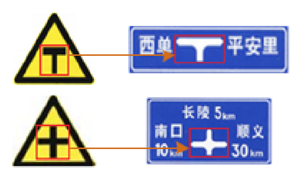

Fig. 12. Similar key features in different traffic signs.

3) Traffic light spotting(TLS)

Due to the color of the traffic light is greatly affected by the natural environment and there are many fake targets detected during the test, such as taillights, street lights or other light sources. Thus, for the training data, we use not only the color information of traffic lights, but also the location information. Some samples of the training data are shown in Fig. 13:

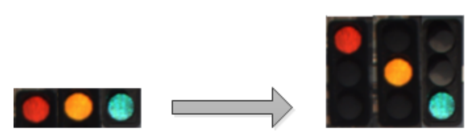

Fig. 13. Samples of traffic lights.

After obtaining the dataset, we get the classification probability of each class directly by Faster-RCNN (Fig. 14).

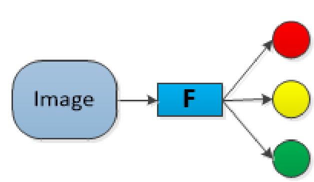

Fig. 14. Traffic light detection and classification framework. F: Faster RCNN framework. Output: the classification probability of each traffic light category.

V. EXPERIMENTAL RESULTS

A. Results and analysis

We implement our method in Tensorflow, Pycharm2017 and openCV3.0.0 in Ubuntu 16.04 with TitanX 12GB memory.

We take three criteria to estimate our method: precision, recall and F-score, which are defined as follows,

$$Precision = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (3)$$

$$Recall = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (4)$$

$$F - measure = \frac{2Precision \times Recall}{Precision + Recall} \quad (5)$$

where n_{TP}, n_{FP} and n_{FN} are the total amount of traffic signal correct detection, the total number of error detection and the total amount of missed detection, respectively.

We not only use a single-model but also use a multi-model. The single-model only contains a detector for all traffic signs, and the model we train using this paper's method is called multi-model. TABLE I is a comparative experiment in TSDR-2017. The TSDR-2017 here is the test dataset which used in the traffic sign recognition task of Intelligent Vehicle Future Challenge 2017, China. As shown in TABLE I, the overall recognition effect of multi-model is better than that of single-model.

TABLE I. THE RESULTS OF THE TWO METHODS IN TSDR-2017

Method	*Precision*	*Recall*	*F-measure*
Single-model	61.7%	47.2%	53.5%
Multi-model	67.3%	58.6%	62.7%

TABLE II. CLASSIFICATION AND STATISTICS OF RESULTS

Class	*Ground-truth*	*Right detect*	*Recognition (%)*
Common signs	960	539 (592)	56.1 (61.7)
Signposts	263	74 (153)	28.1 (58.2)
Traffic lights	120	21 (42)	17.5 (35.0)

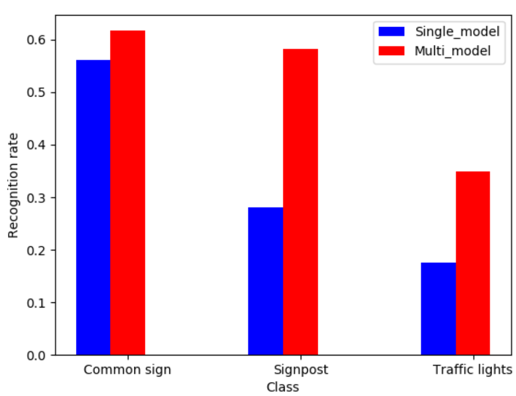

Fig. 15. The comparison of recognition rates between the single-model and the mutil-model on three categories.

TABLE II and Fig. 15 gives the results of two models in common traffic signs, signposts and traffic lights respectively. As expected, the multi-model is superior to the single-model, especially for signposts and traffic lights, and the recognition rate is increased a lot. But our method has a bit unsatisfied performance in traffic lights. The main reason is the number of the training data for traffic lights much smaller than the others.

This paper also does a video-based detection experiment. There are 50 videos in TSDR-2017, each video contains 20 image sequences. The targets in these sequence frames are fixed, and the targets in the later image frames are larger. So we only do detections on the last three frames and we treat the result as the result of the entire video. The experiment result is show in TABLE III, we can see that the video-based detection results have a greater improvement than the image-based detection results.

TABLE III. THE RESULTS OF VIDEO-BASED DETECTION IN TSDR-2017

Method	*Precision*	*Recall*	*F-measure*
Multi-model	82.3%	79.7%	80.9%

Fig. 16. Multiple traffic sign detection results.

Fig. 16 gives some visual effects of our method. It demonstrates that the multi-model can detect various types of traffic signs, and has the ability to detect the icon or text on the signposts and improve the detection performance of small targets.

Fig. 17. Some wrong detection and recognition results.

From the analysis of experimental results, the overall recognition result is not very high, because the picture is shot by a driverless car in the wild. As far as we know, this is the first result testing on the data shot by a driverless car in the wild. In the competition of the traffic sign recognition of Intelligent Vehicle Future Challenge 2017, our method is higher by 1% than the one of Nanjing University of Science and Technology who won the second prize and higher by 3% than the one of Institute of automation of the Chinese Academy of Sciences who won the third prize. In Fig. 17, we can see that error detection is mainly caused by the following reasons: random noise, strong illumination, motion blur and low resolution. Missed detection is mainly caused by the small size and occlusion of the target. Therefore, the detection classification rate can be improved by super-resolution reconstruction, motion deblurring and so on.

VI. CONCLUSIONS

In this paper, we divide traffic signal detection and classification tasks into three sub-tasks, meanwhile, we propose three strategies which based on Faster-RCNN and Inception-CNN to accomplish these sub-tasks. We also build a traffic sign collection which adds the information, facility and service signs and traffic lights based on two competition datasets: TSD-2015 and TSD-2016. The experimental results show that our method has a promising performance in the detection and classification of small targets.

VII. ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61373077.

REFERENCES

[1] Zaklouta F, Stanciulescu B, Hamdoun O. Traffic sign classification using kd trees and random forests[C]//Neural Networks (IJCNN), The 2011 International Joint Conference on. IEEE, 2011: 2151-2155.

[2] Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., & Lopez-Ferreras, F. (2007). Road-sign detection and recognition based on support vector machines. IEEE Transactions on Intelligent Transportation Systems, 8(2), 264-278.

[3] M.Haloi. Anovelplsa based traffic signs classification system. arXiv preprint arXiv:1503.06643, 2015.

[4] T. T. Le, S. T. Tran, S. Mita, and T. D. Nguyen. Real time traffic sign detection using color and shape-based features. In Asian Conference on Intelligent Information and Database Systems, pages 268 – 278. Springer, 2010.

[5] Meuter, M., Nunn, C., Gormer, S. M., Muller-Schneiders, S., & Kummert, A. (2011). A decision fusion and reasoning module for a traffic sign recognition system. IEEE Transactions on Intelligent Transportation Systems, 12(4), 1126-1134.

[6] Ruta A, Li Y, Liu X. Robust class similarity measure for traffic sign recognition [J]. Intelligent Transportation Systems, IEEE Transactions on, 2010, 11(4): 846-855.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1.

[8] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, Jul 2002.

[9] Greenhalgh J, Mirmehdi M. Real-time detection and recognition of road traffic signs [J]. Intelligent Transportation Systems, IEEE Transactions on, 2012, 13(4): 1498-1506.

[10] Ruta A, Li Y, Liu X. Robust class similarity measure for traffic sign recognition [J]. Intelligent Transportation Systems, IEEE Transactions on, 2010, 11(4): 846-855.

[11] Zaklouta F, Stanciulescu B. Real-time traffic-sign recognition using tree classifiers [J]. Intelligent Transportation Systems, IEEE Transactions on, 2012, 13(4): 1507-1514.

[12] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." International Conference on Neural Information Processing Systems Curran Associates Inc. 2012:1097-1105.

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. 770-778.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2818-2826.

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 37(9), 1904.

[16] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.

[17] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, June 1 2017.

[18] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., & Fu, C. Y., et al. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision (pp.21-37). Springer, Cham.

[19] 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), Las Vegas, NV, 2016, pp. 779-788.

[20] Y. Wu, Y. Liu, J. Li, H. Liu, and X. Hu. Traffic sign detection based on convolutional neural networks. In IJCNN, pages 1–7. IEEE, 2013.

[21] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu. Traffic-sign detection and classification in the wild. In CVPR, pages 2110–2118, 2016.

[22] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng and S. Yan, "Perceptual Generative Adversarial Networks for Small Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 2017, pp. 1951-1959.

[23] Namor, A. F. D. D., Shehab, M., Khalife, R., & Abbas, I. (2011). The German Traffic Sign Recognition Benchmark: A multi-class classification competition. International Joint Conference on Neural Networks (Vol.3, pp.1453-1460). IEEE.

[24] Larsson F, Felsberg M. Using Fourier Descriptors and Spatial Models for Traffic Sign Recognition[M]// Image Analysis. Springer Berlin Heidelberg, 2011:238-249.