Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Traffic sign detection method based on Faster R-CNN

To cite this article: Linxiu Wu *et al* 2019 *J. Phys.: Conf. Ser.* **1176** 032045

Traffic sign detection method based on Faster R-CNN

Linxiu Wu[1], Houjie Li[1,*], Jianjun He[1] and Xuan Chen[1]

[1]School of Information and Communication Engineering, Dalian Minzu University, Dalian Liaoning, China

*Corresponding author e-mail: lhj@dlnu.edu.cn

Abstract. Traffic sign detection is the pivotal technology of the traffic sign recognition system. In this article, a traffic sign detection method comes up based on Faster R-CNN deep learning framework. In this method, a convolution neural network is devoted to extract traffic sign image features automatically, and the extracted convolution feature map is sent into a Region Proposal Network (RPN) for foreground objects filtration and regression of bounding boxes. Then the proposed regions are mapped to the feature map, and the fixed-size proposal boxes via Region of Interest pooling layer (RoI). After that, we use the classification network to perform specific classification tasks and further compute the bounding box regression. The experiments performs on the German Traffic Sign Detection Benchmark (GTSDB) and the experimental results show that the method has effectiveness and robustness to different light, block, and motion.

1. Introduction

As artificial intelligence technology advances, the application of autonomous driving technology in vehicles has attracted extensive attention. Traffic sign recognition system is a vital subsystem of the automatic driving system. And Traffic sign detection is the key technology of the traffic sign recognition system. Many researchers have proposed relatively practical traffic sign detection methods that are based on traffic sign color, traffic sign shape and machine learning usually[1]. The color based detection algorithms usually use color space to segment the image. As long as there is less calculation, it can remove numerous non-interested areas, so it has good real-time performance. However, the traffic sign detection based on the image color information is influenced greatly by light and weather [2-8]. The detection method based on the sign shape can also extract the traffic signs effectively from the complex background, Hough Transform is usually used to detect circles and straight lines, and the detection effectiveness is significant. However, a large number of calculations affect the real-time performance, and the single shape makes the detection situation limited (such as triangle, diamond, and circle) [9-12].

Although traffic sign detection method based on machine learning has the advantages of high detection accuracy and timely response to traffic condition changes, however, the characteristics of traditional machine learning depend on manual design, and researchers are required to have considerable experience in the design process, so there are some limitations [13-14].

As deep learning technology advances, the object detection algorithm based on deep learning has been deeply studied, which has received extensive attention in the domain of computer vision. In

IOP Conf. Series: Journal of Physics: Conf. Series **1176** (2019) 032045 doi:10.1088/1742-6596/1176/3/032045

contrast traditional machine learning, deep learning model can automatically extract features without manual design, avoiding the limitation of artificial design features.

This article proposes a method based on the Faster R-CNN deep learning framework to use traffic sign detection. The experiments perform on the GTSDB. It is shown that this method is effective and robust to natural environments such as different lighting, occlusion, motion blur.

2. Faster R-CNN algorithm

Faster R-CNN [15] is a universal target detection algorithm adopt RPN proposed by Ross Girshick's team by R-CNN [16] and Fast R-CNN [17] in 2016. The main idea of the algorithm is to design the RPN network to extract the proposed regions and to generate the proposed regions with the convolution neural network. The convolution layer parameters of the convolutional neural network (CNN) generating the proposed region are shared to the CNN used for classification. This approach makes the algorithm no longer rely on an alone module to generate proposed regions. Then classifying the proposed regions generated and to calculate bounding box regression. Replacing Selective Search with RPN makes the regional proposals time very short, while greatly reducing the time spent on the detection network.

First, the image, category of detection object and object boundary coordinates were input into the Faster R-CNN model. Then we adopt a deep convolutional neural network model to extract the image and take the final convolution layer of the feature map to input the RPN for prediction. The proposed region is mapped to the feature map, the deep features of these regions are extracted from the feature map. The object categories are predicted by Softmax function, to achieve the goals of object classification and boundary regression. The algorithm framework is shown in figure 1 [15].

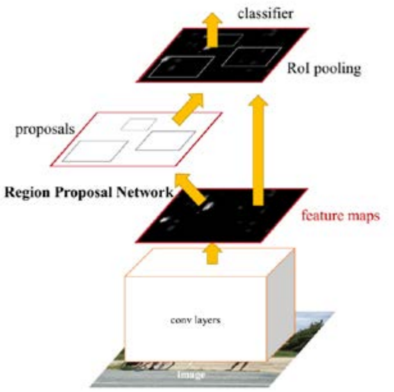

Figure 1. The framework of Faster R-CNN.

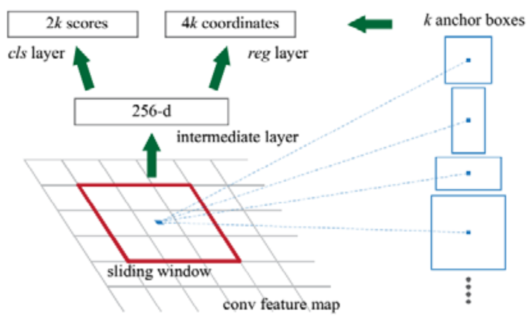

Figure 2. RPN network model.

RPN adopts the method of sliding window selection to generate region proposals on the feature map of the last layer output of the shared convolution network. The input of RPN is an n × n sliding window of the convolution feature map. For each sliding window to predict the object region proposal of k anchors, each anchor has the corresponding scale and proportion. Each point in the convolution feature map is an anchor center, with k corresponding anchors. For a w × h convolution feature map, there are w × h anchors. Each window is passed into the classification network and the regression network simultaneously as a low-dimensional feature vector. The classification network outputs the probability that each anchor belongs to the object. For each window, there are 2k score outputs. The output of the boundary regression network is the translation and zoom value of each anchor, with 4k coordinates output for each window. The RPN network model is shown in figure 2 [15].

3. Traffic sign detection based on Faster R-CNN

Faster R-CNN can achieve the world's advanced object detection rate on PASCAL VOC data set, which can detect 20 types of objects including humans, animals, and vehicles. These objects usually occupy a large proportion in an image, but the object detected in this paper is the traffic sign. The object is small and occupies a small proportion in the image. Therefore, the key parameters need to be modified to achieve good accuracy in the application of traffic sign detection.

3.1. Faster R-CNN parameter modification

In this paper, an algorithm test is carried out for four categories of traffic signs in GTSDB, and the category needs to be set as five categories (including the background). The image size of GTSDB data set is 1360 x 800 pixels, so the input image size in the training and testing phase is changed to 1360 x 800.

In training, the VGG16 network classification model, which is publicly trained on the ImageNet data set, is used as the pre-training model to initialize the parameters of the shared convolutional network layer of RPN network and Fast R-CNN. Set the initial learning rate as 0.001, as the number of iterations increases, the learning rate was reduced by ten times after 50000 iterations, and the training stop for 20,000 iterations.

3.2. Detection algorithm

Step 1. To input the image in the data set to the convolutional neural network (VGG16 network is applied in this article). The CNN propagates forward to the last layer of the shared convolutional layer, inputs the convolution feature map into the RPN, and then propagates forward to the non-shared convolutional layer to obtain the high-dimensional feature.

Step 2. Convolutional feature map generates region proposals and scores of each region through RPN. Non-Maximum Suppression strategy is adopted to pass the proposed regions with score greater than 0.7 and a score less than 0.3 to RoI pooling layer

Step 3. To input RoI pooling layer of proposed regions and high-dimensional feature map for feature extraction.

Step 4. Extracted features into the whole connection layer to the classification score of the regions and bounding box regression as output.

The network structure of this algorithm is shown in figure 3.

In the algorithm, the training process of RPN adopts an end to end mode [18], the loss function is the combined loss of classification error and regression error using the optimization method of back-propagation and stochastic gradient descent, as shown in equation (1).

$$L(\{p_i\}+\{t_i\})=\frac{1}{N_{cls}}\cdot\sum_i L_{cls}\left(p_i,p_i^*\right)+\frac{\lambda}{N_{reg}}\cdot\sum_i p_i^*\cdot L_{reg}\left(t_i,t_i^*\right) \tag{1}$$

Where i represents the i-th anchor point, p_i^* represents the positive sample of the i-th anchor point, and t_i^* represents the deviation between the region proposal and the ground truth. The regression calculation of the anchor box to the adjacent ground truth boundary box is shown in equation (2).

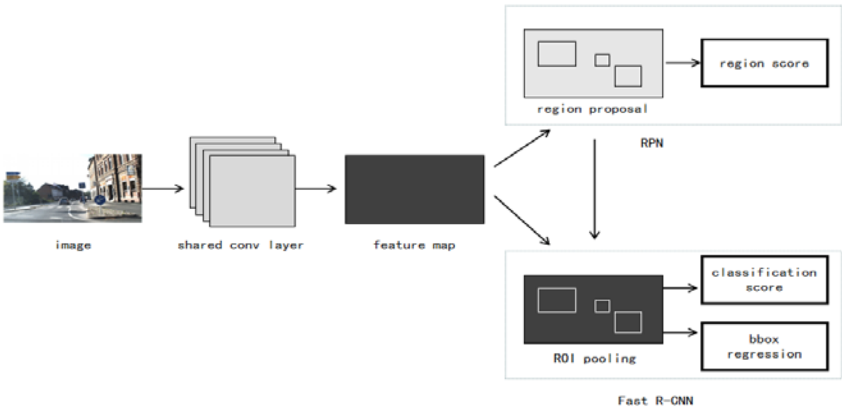

Figure 3. Network structure diagram of traffic sign detection.

IOP Conf. Series: Journal of Physics: Conf. Series **1176** (2019) 032045 doi:10.1088/1742-6596/1176/3/032045

$$\begin{cases} t_x = \dfrac{x - x_a}{w_\alpha}, t_w = \log\left(\dfrac{w}{w_\alpha}\right), t_y = \dfrac{y - y_\alpha}{h_\beta}, t_h = \log\left(\dfrac{h}{h_\alpha}\right) \\ t_x^* = \dfrac{x^* - x_\alpha}{w_\alpha}, t_w^* = \log\left(\dfrac{w^*}{w_\alpha}\right), t_y^* = \dfrac{y^* - y_\alpha}{h_\alpha}, t_h^* = \log\left(\dfrac{h^*}{h_\alpha}\right) \end{cases} \tag{2}$$

x and y express the central coordinates of the proposed region in the formula , while w and h express the width of the region and height of the region. The variable x represents the predicted bounding box, x_α represents the anchor box, and x^* represents the bounding box of the real object (the same is true for variable y, w, and h).

4. Experimental results and analysis

The algorithm in this article trained the Faster R-CNN model on GTSDB. The data set has 900 images, including 600 training images and 300 test images. All the images are 1360×800 pixels in size. There are 1213 traffic signs in the GTSDB, among which 852 are training sets and 361 are test sets. The signs in the image include four categories: prohibitory, danger, mandatory and other.

In this paper, the experiment using the Python language version of the TensorFlow deep learning system in Intel ® Xeon E5-2660 (R) CPU v4 and NVIDIA ® Tesla ® P40 GPU Ubuntu 16.04 system configuration.

We adopt both the average precision (AP) and the mean average precision (mAP) as the evaluation indexes. The AP is the area under PR (Precision-Recall) curve and the mAP is the mean of the AP for all categories. The experimental results are shown in figure 4 that is the PR curve, the mAP is 91.75% that indicates the method in the paper achieves a good performance. Also, the method proposed in this article is robustness for different lighting, motion blur, sign occlusion and so on. Figure 5 (a) - (c) shows the good detection results under three conditions, namely sunny day,weak light day, rainy day. And figure.6 (a) - (c) shows good results in complex natural environments such as motion blur, small size sign and sign occlusion.

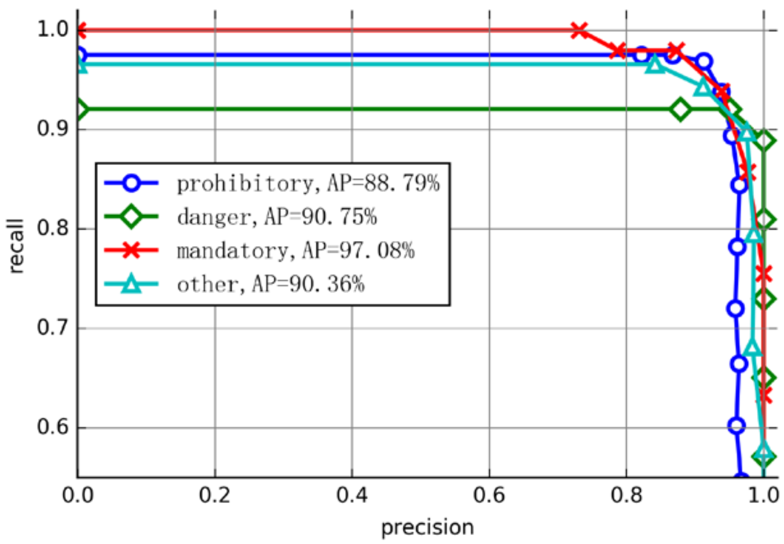

Figure 4. Each category AP results from statistics in the experiment.

(a) Sunny day (b) Weak light (c) Rainy day

Figure 5. Sign detection results under various weather and light conditions.

(a) Motion blur (b) Small size sign (c) Sign occlusion

Figure 6. Sign detection results in special traffic situations.

5. Conclusion

The traffic sign detection algorithm proposed in this article makes use of deep learning technology to automatically extract features and realizes end-to-end training mode by using RPN. The algorithm in this article is carried out on the GTSDB data set, and has achieved a good detection effect, and has good robustness for the detection of signs in complex natural environments such as different lighting, motion blur, sign occlusion and so on.

Acknowledgments

This work was financially supported by the Science and Technology Fund Project of Liaoning Province(20180550273), the Fundamental Research Funds for the Central Universities (No. DC110313, DC201502060405), and the Research Foundation of Education Bureau of Liaoning Province (No. L2014540)

References

[1] Mogelmose A, Trivedi M M, Moeslund T B. Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(4):1484-1497.
[2] Kamada H, Naoi S, Gotoh T. A compact navigation system using image processing and fuzzy control[C]// Southeastcon '90. Proceedings. IEEE. IEEE, 2002:337-342 vol.1.
[3] Janssen R, Ritter W, Stein F, et al. Hybrid Approach For Traffic Sign Recognition[C]// Intelligent Vehicles '93 Symposium. IEEE, 2002:390-395.
[4] Miura J, Kanda T, Shirai Y. An active vision system for real-time traffic sign recognition[C]// Intelligent Transportation Systems, 2000. Proceedings. IEEE, 2002:52-57.
[5] Maldonado-Bascon S, Lafuente-Arroyo S, Gil-Jimenez P, et al. Road-Sign Detection and Recognition Based on Support Vector Machines[J]. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(2):264-278.
[6] Nguwi Y Y, Kouzani A Z. Detection and classification of road signs in natural environments[J]. Neural Computing & Applications, 2008, 17(3):265-289.
[7] Xu S. Robust traffic sign shape recognition using geometric matching[J]. Iet Intelligent Transport Systems, 2009, 3(1):10-18.
[8] Wang G, Ren G, Jiang L, et al. Hole-based traffic sign detection method for traffic signs with red rim[J]. Visual Computer, 2014, 30(5):539-551.

[9] Barnes N, Zelinsky A, Fletcher L S. Real-Time Speed Sign Detection Using the Radial Symmetry Detector[M]. IEEE Press, 2016.
[10] Meuter M, Nunn C, Gormer S M, et al. A Decision Fusion and Reasoning Module for a Traffic Sign Recognition System[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4):1126-1134.
[11] Álvaro Gonzalez, Garrido M Á, Llorca D F, et al. Automatic Traffic Signs and Panels Inspection System Using Computer Vision[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2):485-499.
[12] Creusen I I, Wijnhoven R R, Herbschleb E, et al. Color exploitation in HOG-based traffic sign detection[J]. 2010.
[13] Qian R, Zhang B, Yue Y, et al. Robust chinese traffic sign detection and recognition with deep convolutional neural network[C]// International Conference on Natural Computation. IEEE, 2016:791-796.
[14] Mukhometzianov R, Wang Y. Review. Machine learning techniques for traffic sign detection[J]. 2017.
[15] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137-1149.
[16] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. Computer Vision and Pattern Recognition. IEEE, 2013: 580-587.
[17] Girshick R. Fast R-CNN[J]. Computer Science, 2015.
[18] .Jiao L, Zhao J, Yang S, et al. Deep Learning, Optimization and Recognition[M]. Tsinghua University Press, Beijing, 2017.