World Scientific Research Journal Volume 6 Issue 12, 2020
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

# Identification and Detection of Traffic Signs based on Faster R-CNN

Ao Qin[1], Yujun Zhang[1]

[1]University of Science and Technology Liaoning, Anshan, China

## Abstract

**This paper applies the Tensorflow deep learning framework and adopts the target detection algorithm in deep learning to solve the problem of road traffic sign identification. First, 90.8%mAP was obtained by training the deep learning target detection algorithm of Faster R-CNN, and then 93.12%mAP was obtained by training the data set into large target, medium target and small target.**

## Keywords

**Target detection algorithm; Deep learning; Faster R - CNN; Sub-target training; mAP.**

## 1. INTRODUCTION

The field of intelligent transportation system has gradually developed and become a research hotspot [1]. Traffic sign recognition system is an important part of it [2]. Traffic signs have specific shape and color information, so their detection and recognition is a restricted problem [3]. The main tasks of the traffic sign recognition system can be divided into two stages [4]: the detection of traffic signs, which mainly includes image acquisition, pre-processing image and threshold segmentation; Secondly, the identification of traffic signs mainly includes the extraction of traffic signs features and the classification of traffic signs. The research on traffic sign detection and identification based on Faster R-CNN is of great significance to vehicle autonomous driving and unmanned driving.

## 2. DEEP LEANRING FRAMEWORK

### 2.1. Faster R-CNN

Ross B. Girshick proposed a new Faster R-CNN in 2016[5]. Faster R-CNN mainly formed a breakthrough in detection and identification process and overall structure.

On the process: Faster R-CNN first feeds the whole picture into the CNN neural network to obtain the feature map, and then inputs the obtained feature map to RPN to obtain the feature information of the candidate box. Then, according to the features extracted from the candidate box, the classifier is used to determine which particular class it belongs to and which feature of the candidate box. Finally, the regression is used to further adjust the box to a more accurate position.

Structurally: Faster R-CNN has integrated feature extraction, proposal extraction, bounding Box regression, and classification into a network, which greatly improves the comprehensive performance and greatly improves the detection speed.

### 2.2. Network Structure Analysis of Faster R-CNN

Faster RCNN can be divided into four main contents:

World Scientific Research Journal Volume 6 Issue 12, 2020
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

(1) the Conv the layers. For example, VGG16, a CNN network target detection method, and Faster RCNN first uses VGG16 convoluted network to extract feature maps of image. feature maps are shared for subsequent RPN layer and full connection layer.

(2) the Region Proposal Networks. The RPN network is used to generate region proposals. Evaluation anchors belong to positive or negative proposals through Softmax, and evaluation Box Regression is used to correct them to get precise proposals.

(3) the Roi Pooling. The input feature maps and Proposals collected by this part of the network are combined to extract proposal feature maps and send them to the subsequent full connection layer to determine the target categories.

(4) Classification. The category of the proposal is calculated using proposal feature maps, and the final precise position of the test box is obtained again with bounding Box regression. The network structure of Faster RCNN is shown in Figure 1.

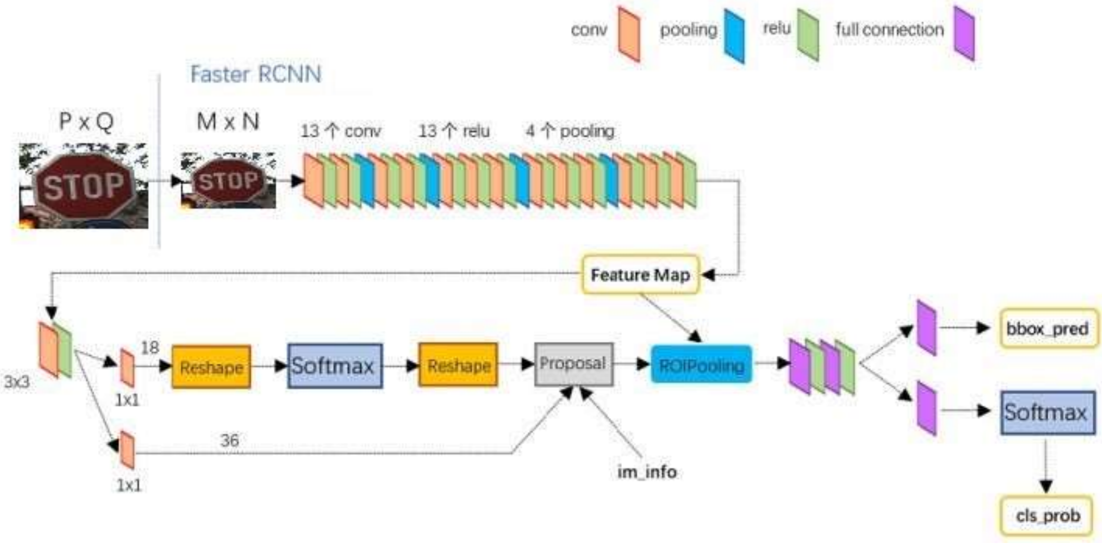


**Figure 1.** Faster R-CNN network structure

## 2.3. The Core Structure of Faster R-CNN

2.3.1. VGG16

VGG16 consists of 13 Convolutional layers, 3 Fully connected layers and 5 Pool layers. Among them, the Convolutional Layer and the Fully connected Layer have weight coefficients, so they are also called weight layers. As shown in Figure 2, the VGG16 network in the Faster R-CNN of this paper contains 13 CONv layers, 13 relu layers, and 4 pooling layers. It's different from the basic VGG16.

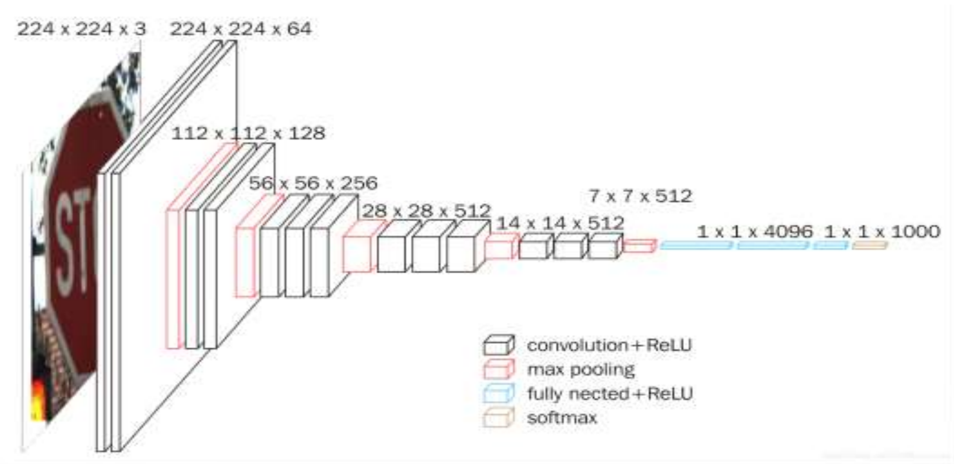


**Figure 2.** VGG16 Network structure

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

**Table 1.** Deep learning device diagram

| type | Filter | Size/Stride | output |
|---|---|---|---|
| convolution | 64 | 3×3/1 | 224×224×64 |
| max pooling | | 2×2/2 | 112×112×64 |
| convolution | 128 | 3×3/1 | 112×112×128 |
| max pooling | | 2×2/2 | 56×56×128 |
| convolution | 256 | 3×3/1 | 56×56×256 |
| max pooling | | 2×2/2 | 28×28×256 |
| convolution | 512 | 3×3/1 | 28×28×512 |
| max pooling | | 2×2/2 | 14×14×512 |
| convolution | 512 | 3×3/1 | 14×14×512 |
| max pooling | | 2×2/2 | 7×7×512 |

As shown in Table 1 above, the input image size is 224×224×3, the 3×3 convolution kernel with 64 channels is 3, step size is 1, padding=same, convolution is applied twice, and then activated by ReLU, the output size is 224×224×64. After Max pooling, the filter is 2×2, the step length is 2, the image size is halved, and the pooling size becomes 112×112×64. After 128 3×3 convolution kernels and twice convolution, ReLU is activated and the size becomes 112×112×128. Max pooling, size of 56×56×128. After 256 3×3 convolution kernels and three times of convolution, ReLU is activated and the size becomes 56×56×256. Max pooling and the size becomes 28×28×256. After 512 3×3 convolution kernels and three times of convolution, ReLU is activated and the size becomes 28×28×512. Max pooling, size 14×14×512. After 512 3×3 convolution kernels, the ReLU is convolved with three times, and the size becomes 14×14×512. Max pooling, size 7×7×512.

2.3.2.RPN

The core idea of RPN is to use CNN convolutional neural network to directly produce regional proposals, and the method used is to slide through the final convolutional layer, because the anchor mechanism and border regression can obtain regional proposals with multiple scales and multiple length-width ratios.

RPN network is a full convolutional network, which can be trained end-to-end for the task of generating detection suggestion box, and can predict the boundary and score of object at the same time. RPN adds 2 additional convolutional layers on CNN. In this paper, RPN network USES VGG16 for feature extraction, and the composition form of RPN network can be expressed as VGG16+RPN.

The specific process of RPN is as follows: slide on the feature graph after CNN convolution, the sliding network is fully connected to the window on the feature graph each time, and then maps to a low-dimensional vector, which is finally sent to the two full connection layers, bbox regression layer and Box classification layer. The bbox regression layer predicts the coordinates of the proposal anchor corresponding to the proposal, and the Box classification layer judges whether the proposal is the background or the detection target, the structure of RPN network is shown in Figure 3.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

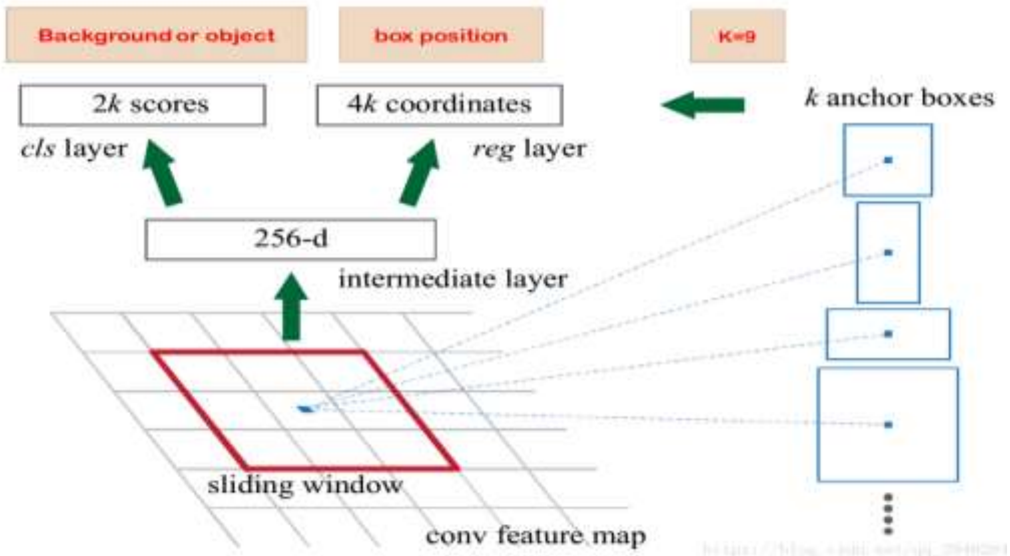


**Figure 3.** RPN network structure

## 2.4. Small Target

Introduction to small targets: There are two definitions: one is the relative size, for example, the length and width of the target size is 0.1 of the original image size, which can be considered as a small target; the other is the definition of absolute size, that is, a target with a size less than 32*32 pixels can be considered as a small target. According to the second definition mode of small targets, the data set is divided into large, medium and small targets for detection and identification, and compared with the original classification. The classification results are shown in Table 2.

**Table 2.** By target

| Title | ClassName | Min rectangle | Max rectangle |
|---|---|---|---|
| Small Object | 0 | 0×0 | 32×32 |
| Medium Object | 1 | 32×32 | 96×96 |
| Large Object | 2 | 96×96 | ∞×∞ |

## 2.5. German Traffic Dataset

The training set contains 39,209 pictures of traffic signs, labeled as 43 categories. The number distribution of pictures in each category is shown in Table 4. It can be seen that the data distribution of the data set of German traffic signs is uneven. There are also some problems in the data set, such as overexposure, too dark pictures and unclear images.

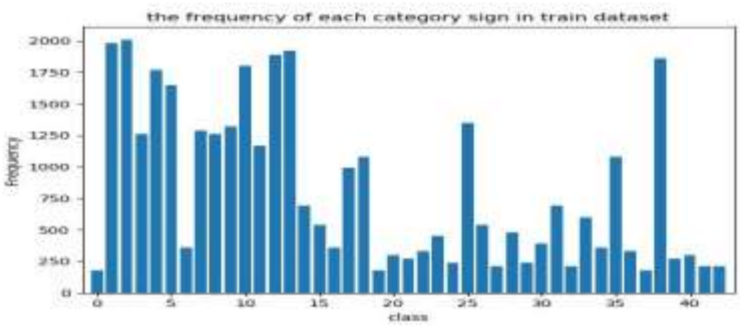


**Figure 4.** Number distribution of 43

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

## 3. THE EXPERIMENT

The experimental equipment uses the deep learning computer in the laboratory. The specific parameters of the computer are shown in Table 3.

**Table 3.** Deep learning device diagram

| equipment | The parameter value |
|---|---|
| The graphics card | GeForce RTX 2070 SUPER(8G) |
| The processor | i9-9900k CPU @3.60GHZ 3.60GHZ |
| memory | 64 GB |
| system | Windows 10 64 bit |

### 3.1. Experimental Results and Analysis

The parameters of Faster R-CNN in this paper are shown in Table 4.

**Table 4.** Experimental parameters

| Parameter names | The parameter value |
|---|---|
| BATCH_SIZE | 256 |
| MAX_ITERS | 120000 |
| STEP_SIZE | 10000 |
| LEARNING_RATE | 0.0003 |
| MOMENTUM | 0.9 |
| SNAPSHOT_ITERATIONS | 40000 |

Evaluation standard: mAP

First, the Average Precision (AP) of each category is calculated, and then the Average Precision of all categories is obtained.

A picture a category target number a (Total objects), the number of correctly predicted for b (True Positives), then the model for the accurate rate of the class P (Precision) is b/a, such as formula (1) according to the number of data sets, each image has the accurate rate of this category, the category of the accurate rate averaged all images, the average is the average accuracy of the class, such as formula (2). Using these Average Precision values to judge the performance of the model for any given category, the data set generally has multiple categories, and the Average Precision of each category can be calculated, and the Mean Average Precision of each category can be calculated by adding the Average Precision of each category and calculated its average, which is the evaluation standard mAP (as shown in Formula (3)).

$$\mathrm{p} = \frac{\text{True positives}}{\text{Total objects}} \tag{1}$$

$$\mathrm{AP} = \frac{\sum p}{\text{Total images}} \tag{2}$$

$$\mathrm{mAP} = \frac{\sum AP}{\text{Total classcs}} \tag{3}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

Experimental analysis: In this experiment, 90.8%mAP was obtained by applying the German traffic sign data set according to the original 43 classification training, and then 93.1%mAP was obtained after the original data set was reclassified according to the large, medium and small targets. Among them, Figure 5 is the Precision -recall curve of the classification of small targets, the experimental results are shown in Table 5.

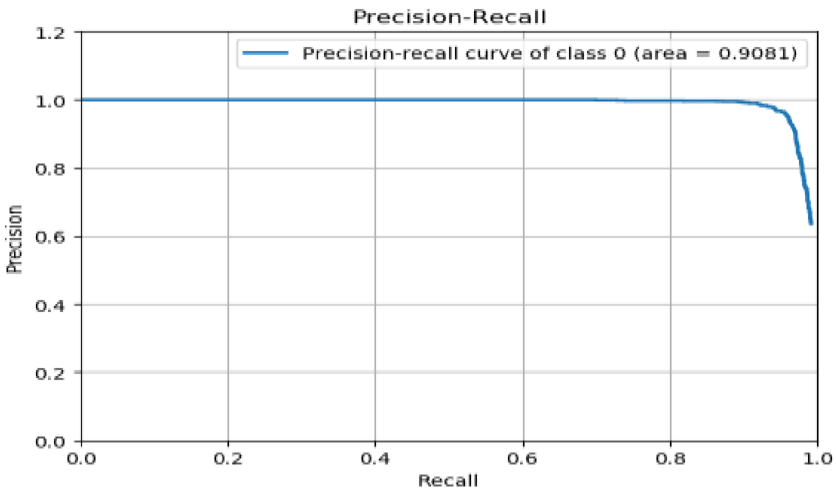


**Figure 5.** Small target Precision – Recall

**Table 5.** The experimental results

| Experiment | mAP |
|---|---|
| Faster-RCNN vgg16 | 90.8 |
| Classification Faster-RCNN vgg16 | 93.1 |

## 4. CONCLUSION

The paper analyzes the detection and identification process of THE Faster R-CNN deep learning network, analyzes the RPN network and the VGG16 network, and takes the German traffic sign as the research object. The original 43 classification reaches 90.8mAP, and the large, medium and small targets are reclassified, and the training reaches 93.1%mAP. By classifying different targets, the ability of the model to deal with different targets can be better seen, laying a foundation for future research. Meanwhile, research on traffic sign detection and identification is a hot spot at present, promoting the development and progress of unmanned driving and other practical fields.

## ACKNOWLEDGMENTS

I would like to extend my deep gratitude to Professor Yu Jun Zhang who have offered me a lot of help and support in the process of my thesis writing.

## REFERENCES

[1] HUANG H,HOU LY.Speed limit sign detection based on Gaussian color model and template matching[c].IEEE International Conference on Vision,Image and Sig-nal Processing,2017: 118−122.

[2] XU Z,REN J Y,BAO C Q. A Traffic signs' detection method of contour approximation based on concave re-moval[c]. 2016 Chinese Control and Decision Confer-ence,2016: 5199−5204.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202012_6(12).0037

[3] YANG Y,LUO H L,XU HR,et al. Towards real −time traffic sign detection and classification[J].. IEEE Transactions on Intelligent Transportation Systems,2016,17(7) : 2022−2031.

[4] WANG C Y.Research and application of traffic sign de-tection and recognition based on deep learning[c]. In-ternational Conference on Robots &Intelligent System,2018: 150−152.

[5] Ren S,He K, Girshick R,etal. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.[J].IEEE Transactions on Pattern Analysis&Machine Intelligence,2017, 39(6):1137-1149.