A Small Traffic Sign Detection Algorithm Based on Modified SSD

Hongxin Shan and Wenxing Zhu*
Shandong University, Jinan, China
*zhuwenxing@sdu.edu.cn

Abstract. Traffic sign detection is an important part of many systems such as autonomous driving, driver safety and assistance. In this paper, the detection capability of SSD for small targets is analyzed and improved based on ssd_300 model. CSUST Chinese Traffic Sign Detection Benchmark (CCTSDB) dataset is used to train the model for Chinese road traffic conditions. The improved model was compared with ssd_300 model. The experimental results show that the mAP of the improved model on the test dataset achieves 0.85, which is 0.13 higher than ssd_300, and the algorithm can reach real-time detection. The improved model can effectively detect three categories of Chinese traffic signs and has strong robustness against various disturbances.

1. Introduction
To aid safe driving and effective navigation, traffic signs provide essential traffic information that is an important part of autonomous driving, driver safety assistance and traffic monitoring systems[1]. Traffic sign detection is a challenging task due to the complexity of the road scene, Light brightness change, occlusion of the object, and the relatively small relative area of the detection object. Therefore, traffic sign detection task has received much attention, and a large number of researchers have conducted in-depth research on it. Liang M et al, present a traffic sign detection model consisting of two modules [2], Extracting regions of interest (ROI) through supervised learning, then detects traffic signs using a combination of Histograms of Oriented Gradients (HOG) and Support Vector Machines (SVM). Zhu YY et al[3], propose two methods of combining neural networks, using fully convolutional network (FCN) to guid traffic sign proposals and deep convolutional neural network (CNN) for object classification, in 2016.

In recent years, many advanced object detection algorithms have achieved huge success, such as R-FCN[4], Faster R-CNN[5], SSD[6], YOLO[7], and so on. Among them, the SSD object detection algorithm has a good balance between the detection accuracy and the detection speed.

2. Related work

2.1. Analysis of small target detection by SSD
In SSD, the pixel above each feature map corresponds to several default anchor box. This network trains the anchor to drive the feature training. Select the IOU (Intersection Over Union) of prediction box larger than 0.5 as the positive sample, and the prediction box with the IOU less than 0.5 as the negative sample. Therefore, there are many predicted bounding box with the IOU more than 0.5 on the large target object, the positive and negative samples are balanced. On the small target, there are fewer
predicted bounding box with the IOU greater than 0.5, which leads to the imbalance of the number of positive and negative objects in the small target object, which is not conducive to the training of small target objects.

On the other hand, SSD, YOLO and other single-stage multi-scale algorithms, small target detection requires higher resolution, so it is mainly detected at the bottom feature layer, such as conv4_3 in SSD, but the semantic information of the bottom feature layer is not rich enough. For the high-level feature layer, the feature extraction content is more abstract. The more features are extracted, the less corresponding fine-grained information, and less sensitive to the detection of small targets. The size of each feature layer of SSD_300 and the size of the corresponding default anchor box are shown in Table 1.

| Layer  | Feature shape | Anchor size bound | Anchor size |
|--------|---------------|--------------------|-------------|
| Conv4_3 | 38×38         | 0.07               | 21×21       |
| Conv7  | 19×19         | 0.15               | 45×45       |
| Conv8_2| 10×10         | 0.33               | 99×99       |
| Conv9_2| 5×5           | 0.51               | 153×153     |
| Conv10_2| 3×3          | 0.69               | 207×207     |
| Conv11_2| 1×1          | 0.87               | 261×261     |

It can also be seen from Table 1 that the anchor size (Mapping Region scale) of the bottom feature layer is smaller, while that of the upper layer is larger. In summary, in order to improve the detection ability of small object, it is necessary to set an appropriate anchor size and increase the semantic information of the bottom feature layer[8].

2.2. Dataset

Most of the previous research on traffic sign detection algorithm is based on the German traffic sign detection benchmark (GTSDB)[9]. There are many reasons, on the one hand it is widely accepted to compare the traffic sign detection methods in literature, on the other hand a special public challenge has been set for it and a large number of participants have contributed the results. The GTSDB divides traffic signs into three categories: mandatory signs, danger signs and prohibited signs, which contain the natural traffic scenes of various types of roads under different lighting conditions. The data set was composed of 900 images with a size of 1360x800. It was divided into a training set consisting of 600 images (846 symbols) and a test set consisting of 300 images (360 symbols).

In order to conduct research on the detection of Chinese traffic signs, a data set for Chinese traffic conditions was selected. This paper focuses its experimentation on the CSUST Chinese Traffic Sign Detection Benchmark dataset (CCTSDB)[10]. Similar to the GTSDB, CCTSDB classifies Chinese traffic signs into three categories: mandatory signs with blue color and circular or rectangular shape, warning signs with yellow color and triangular shape, and prohibitory signs with red color and circular shape. The open source CCTSDB has a total of 15,723 images, including different types of road (highway, street, rural and urban) under the condition of different illumination of natural traffic scene. The picture of the size of 1000 x350 accounted for 77% of the total data set, the size of the rest of the images (1024×768, 1280×720, 912×684, 513×999, 641×936, 457×889, 852×538, 1022×504, 759×497) accounted for 33% of the total data set. Because the open source CCTSDB data set did not divide the training dataset and the test dataset, in this paper, 12,516 images were used as the training dataset, and 1000 images were evenly extracted from all the data sets as the test dataset. It is worth noting in this
article, make a small amount of clipping of the width of partly 1000×350 image in the test dataset. The distribution of traffic sign types in the training dataset and the test dataset is shown in Table 2.

Table 2. Distribution of traffic sign types in the training dataset and the test dataset

| Types of sign | Training dataset | Test dataset |
|---------------|------------------|--------------|
|               | Picture number   | Sign number  |
|               |                  | Picture number | Sign number |
| warning       | 3138             | 3569          | 239         | 283         |
| prohibitory   | 6569             | 9503          | 534         | 734         |
| mandatory     | 4644             | 6064          | 345         | 454         |
| total         | 12516            | 19136         | 1000        | 1471        |

In the natural road scenes, the detection of traffic signs is often the detection of small object. In the object detection algorithm, there are two ways to define small object, one is the relative size, for example, the length and width of the object is less than 0.1 of the original image size, which can be considered as a small object; the other is the definition of absolute size, the object smaller than 32*32 pixels can be considered as a small object. The relative size of the CCTSDB dataset ground truth is shown in Figure 1.

Figure 1. Relative size of the CCTSDB dataset ground truth

It can be seen from Figure 1 that the relative size of most ground truth in CCTSDB dataset is less than 0.1, therefore, most ground truth of CCTSDB dataset is a small object. It can also be seen that the aspect ratio of most ground truth is not 1:1. In this paper, the anchor size of the first feature map is set to 10, and the number of default box is set to 6, and the anchor ratios are: 1:1*2,1:2,2:1,1:3,3:1.

3. Traffic sign model

3.1. Architecture

According to the SSD detection of small object and the description of CCTSDB dataset in the second section, in order to improve the detection effect of the model on small object, high-level features map and low-level features map are fused. In our model, the same size as conv4_3 is obtained by deconvolving conv5_3. The following, convolution and Relu activation operations are performed on conv4_3 and the deconvolved conv5_3. Then we merge them and perform a convolution operation to get a new feature map. Similarly, conv7 and block9/conv1×1(size:19×19) perform the same operation for feature fusion. The fusion of high level features and low level features can obtain the characteristics of the receptive field without lacking semantic information. Because the semantic information will affect the detector's determination of whether the detection area is an object or background, feature fusion of the front feature layer can enhance the ability to detect small targets.
It is worth mentioning that there is no large ground truth in the CCTSDB dataset, and it can be seen from Figure 1 that the relative size of the largest object does not exceed 0.6. It can be seen from Table 1 that the anchor size of conv11_2 feature map is much larger than the relative size of the largest object. Therefore, the last feature map (block11) was removed from our model without affecting the detection effect. The improved model architecture is shown in Figure 2.

Figure 2. The architecture of our model

3.2. Training
In this paper, two models are trained. One is the original ssd_300 model and the other is our improved model. In order to reduce the time of training model, the transfer learning was adopted to train the model. Through transfer learning, the CCTSDB dataset was used to fine-tune the model to detect and classify the traffic signs based on shape and color: warning(1), prohibitory(2) and mandatory(3). During the training, the public vgg16 weight was used to initialize the weight of our network for feature extraction (conv1_x, conv2_x, conv3_x, conv4_x, conv5_x) to train the weight of the remaining feature map until convergence. Next, the trained model weight is used to initialize the network to train the complete network until convergence. The parameters of the train network is shown in Table 3.

Table 3. Parameter Settings

| Parameter         | Value   |
|-------------------|---------|
| Learning rate     | 0.001   |
| Optimizer         | adam    |
| Batch size        | 24      |
| Weight decay      | 0.0005  |
| GPU memory fraction | 0.9    |

4. Results
In this paper, ssd_300 is compared with our improved model. In our experimental evaluation of the model, the training and testing were carried out on a Linux PC, and the machine configuration was described in part B of the section III. The test dataset contains 1000 pictures and 1471 ground truth. The Mean Average Precision (mAP) and the Frame per Second (FPS) of the two models are shown in Table 4.

Table 4. FPS and mAP for each model

| Model  | mAP | FPS | AP of warning | AP of prohibitory | AP of mandatory |
|--------|-----|-----|---------------|--------------------|-----------------|
| ssd_300 | 0.715 | 28  | 0.694         | 0.734              | 0.717           |
| ours    | 0.85 | 25  | 0.859         | 0.845              | 0.846           |
It can be seen from Table 4 that the detection ability of the improved model for traffic signs is greatly improved compared with the model before the improvement (ssd_300). The mAP achieves 0.85, which is 0.13 higher than the model before the improvement. The detection speed on the experimental platform can achieve real-time detection. The precision-recall curves (PR Curve) of each traffic sign categories are shown in Figure 3.

![PR Curve for each categories](image)

(a). The pr curve of warning

(b). The pr curve of prohibitory

(c). The pr curve of mandatory

**Figure 3. PR Curve for each categories**

The experimental results under different scenes are shown in Figure 4. It can be seen from Figure 4 that our model can effectively detect three categories of Chinese traffic signs and has strong robustness against various disturbances.
5. Conclusion
A traffic sign detection algorithm based on modified SSD is proposed in this paper. This method uses CCTSDB dataset training model for Chinese traffic signs. The improved model was compared with ssd_300 model. The experimental results show that the mAP of the improved model on the test dataset achieves 0.85, which is 0.13 higher than ssd_300, and the detection can reach real-time detection, but the confidence of our model on the small traffic signs is low.

In the future, we will continue to study the detection of small traffic signs to improve the confidence of detection.

References
[1] De L, Moreno L, Salichs M and Armingol J 1997 Road traffic sign detection and classification. IEEE Transactions on Industrial Electronics 44(6) pp 848-859.
[2] Liang M, Yuan M, Hu X, Li J and Liu H 2013 Traffic sign detection by ROI extraction and histogram features-based recognition. In The 2013 international joint conference on Neural networks (IJCNN) pp 1-8. IEEE.
[3] Zhu Y, Zhang C, Zhou D, Wang X, Bai X and Liu W 2016 Traffic sign detection and recognition using fully convolutional network guided proposals. Neurocomputing pp 758-766.
[4] Dai J, Li Y, He K and Sun J 2016 R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems pp 379-387.
[5] Ren S, He K, Girshick R and Sun J 2015 Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems pp 91-99.
[6] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C 2016 Ssd: Single shot multibox detector. In European conference on computer vision pp 21-37.
[7] Redmon J and Farhadi A 2017 YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition pp 7263-7271.
[8] Jeong J, Park H and Kwak N 2017 Enhancement of SSD by concatenating feature maps for object detection. In BMVC.
[9] Houben S, Stallkamp J, Salmen J, Schlipsing M and Igel C 2013 Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In The 2013 international joint conference on neural networks (IJCNN) pp 1-8. IEEE.
[10] Zhang J, Huang M, Jin X and Li X 2017 A real-time chinese traffic sign detection algorithm based on modified YOLOv2. Algorithms 10(4) p 127.