

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

# An Improved Light-weight Traffic Sign Recognition Algorithm Based on YOLOv4-Tiny

LANMEI WANG<sup>1</sup>, KUN ZHOU<sup>1</sup>, ANLIANG CHU<sup>1</sup>, GUIBAO WANG<sup>2</sup>, LIZHE WANG<sup>1</sup>

<sup>1</sup> School of Physics and Optoelectronic Engineering, Xidian University, Xi'an, 710071, China

<sup>2</sup> School of Physics and Telecommunication Engineering, Shaanxi University of Technology, Hanzhong, 723001, China

Corresponding author: Lanmei Wang (e-mail: lmwang@mail.xidian.edu.cn).

This work is supported by the National Natural Science Foundation of China under Contract [Grant No. 61772398, 61972239]; the Key Research and Development Program Projects of Shaanxi Province [Grant No. 2019SF-257, 2020GY-024, 2021GY-182].

**ABSTRACT** Aiming at the problems of low detection accuracy and inaccurate positioning accuracy of light-weight network in traffic sign recognition task, an improved light-weight traffic sign recognition algorithm based on YOLOv4-Tiny was proposed. By improving the K-means clustering algorithm, the anchor with appropriate size is generated for the traffic sign data set to improve the detection recall rate and target positioning accuracy. The strategy of large-scale feature map optimization is proposed, which enriches the feature level of the network by using the low-level information, strengthens the representation of the feature information of the small target, and improves the detection accuracy of the long-range small target. In view of the problem of missed detection of high overlapping targets in the post-processing stage of the model, the paper proposes an improved NMS algorithm to screen the prediction box, avoid deleting the prediction results of different targets, and further improve the detection accuracy and recall rate of the target. Experimental results show that, compared with the original YOLOv4-Tiny algorithm, the improved algorithm in traffic sign recognition task based on TT100K dataset, mAP and recall are improved by 5.73% and 7.29% respectively, and FPS value is maintained at about 87 f/s, which meets the accuracy and real-time requirements of traffic sign recognition task.

**INDEX TERMS** traffic sign recognition, YOLOv4-Tiny, clustering algorithm, large scale, improved NMS algorithm

## I. INTRODUCTION

The identification of traffic signs is an important research content in the field of automobile autonomous driving system and driver-assisted driving system. Traffic signs contain a lot of useful information, which can prompt drivers to make a correct response to road condition information in real time, greatly reduce the occurrence of traffic accidents and improves the safety of driving [1]. Therefore, the study of fast and accurate traffic sign recognition system under the real scene has important practical value and a wide range of application scenarios.

### A. Literature Review

Traffic sign recognition is often carried out in a complex outdoor environment, which is vulnerable to the interference of natural environment and human factors, such as bad weather, different light and shade, sign blocking and damage, etc., thus causing recognition difficulties [2]. Based on this, in the research of traffic sign recognition, a large number of

complex algorithms have been proposed [3]. Traditional algorithms mainly rely on artificial feature extraction, such as local binary pattern (LBP) [4], Gabor [5], histogram of oriented gradient (HOG) [6], etc., and use support vector machine (SVM) [7], AdaBoost [8] and other classifiers to complete traffic sign recognition. However, in the face of complex outdoor environment, artificial feature extraction can not meet the actual needs.

In recent years, deep learning model has gradually become the main algorithm in the field of target detection, and convolutional neural network has achieved remarkable results in the field of target detection [9-10]. The target detection model based on deep learning can be divided into two types: a two-stage detection algorithm based on region proposal, such as region convolutional neural networks (R-CNN) [11], Fast R-CNN [12], Faster R-CNN [13], etc. This two-stage detection algorithm needs to generate the target candidate box first, and then classify and regress the target candidate box. The other is single-stage detection

algorithms such as you only look once (YOLO) [14-17], single shot multibox detector (SSD) [18], which can directly extract features from the network to predict object classification and location. In contrast, the two-stage detection method has high precision but slow speed, and the single-stage detection algorithm is fast but less accurate.

At present, many scholars apply target detection algorithm to traffic sign detection. Zuo [19] et al use Faster R-CNN to detect traffic signs. You [20] et al tailor-made the network to reduce the computational complexity based on SSD algorithm and applied it to traffic sign detection. Yang [21] and yuan [22] et al extract the region of interest from the input image by adding an attention module to the convolutional neural network to refine the feature extraction of traffic signs under complex background. Zhang [23] effectively use fine-grained features at the bottom to achieve accurate target positioning through image enhancement and the introduction of spatial pyramid pooling (SPP) module in YOLOv3.

### B. Main Work

Although the existing deep learning methods have achieved some results in the task of traffic sign detection, they still have limitations in the face of complex natural environment. For example, the traffic sign algorithm model for real-time detection has low recall and detection accuracy. The influence of shooting angle makes the traffic signs overlap, resulting in missing detection of some targets. Therefore, in order to solve the above problems, this paper proposes several improvement strategies based on the lightweight version YOLOv4-Tiny algorithm of YOLOv4,

so as to improve the detection accuracy and the robustness of the model on the premise of meeting the real-time performance. The main works are as follows:

(1) The K-means clustering algorithm is improved to generate anchor boxes. By introducing generalized IOU (GIoU) [24] instead of intersection over union (IOU) in the distance calculation formula in the K-means clustering algorithm, GIoU can consider the area of non-overlapping areas, that is, when the two boxes do not completely intersect, GIoU introduces non-overlapping area items, which can better reflect shape information, thus improving the recall rate of the algorithm model and accelerating the convergence of the model.

(2) A large-scale feature map optimization strategy is proposed, and the size of the two-scale output feature map of YOLOv4-Tiny is changed from the original  $19 \times 19$  and  $38 \times 38$  to  $38 \times 38$  and  $76 \times 76$ , so as to improve the detection accuracy of small targets.

(3) An improved algorithm based on soft non-maximum suppression (NMS) [25] is proposed. Aiming at the problem of missing detection caused by overlapping traffic signs in the traffic sign data set, the missing detection rate is reduced and the generalization of the model is improved by improving the score reset function.

## II. ALGORITHM MODEL

YOLOv4-Tiny is a simplified version of YOLOv4. Compared with YOLOv4, YOLOv4-Tiny has a faster detection speed, but the accuracy has declined.

Fig. 1 shows the network structure diagram of YOLOv4-

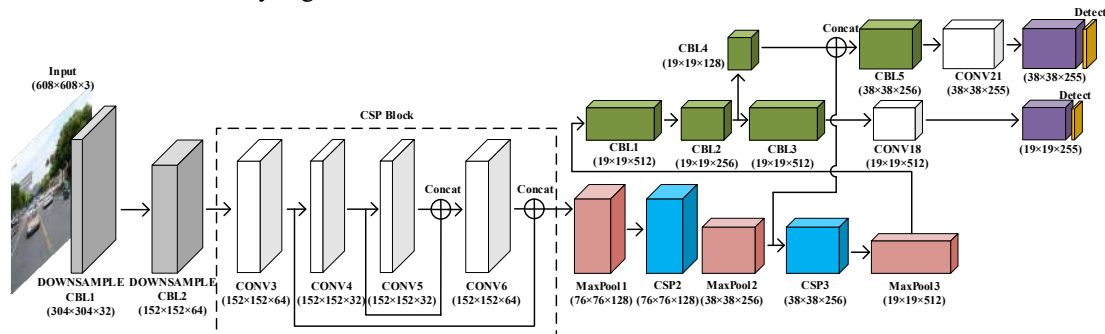


FIGURE 1. YOLOv4-Tiny network structure

Tiny. Compared with YOLOv4, the backbone network of YOLOv4-Tiny is greatly simplified. Feature pyramid network is used for 32 times down-sampling and 16 times down-sampling to get two kinds of different sizes of feature map for target detection, which improves the detection speed.

There are several basic components in the YOLOv4-Tiny network structure. The CBL module consists of Conv convolution, BN normalization and Leaky-Relu activation function. The implementation method of down-sampling CBL module is that the stride length of convolution kernel is set as 2 to achieve the purpose of down-sampling. CSP

module uses CSPNet network structure for reference and is composed of CBL module and Concat tensor splicing module to better integrate feature information.

Assuming the input image size is  $608 \times 608 \times 3$ , the details of the backbone network parameters of YOLOv4-Tiny are shown in Table I.

TABLE I  
YOLOV4-TINY BACKBONE NETWORK PARAMETERS

Network layer	Kernel Size	Stride	Feature map
Input	—	—	$608 \times 608 \times 3$
CBL1	$3 \times 3$	2	$304 \times 304 \times 32$

CBL2	3×3	2	152×152×64
CSP1	3×3/1×1	1	152×152×64
MaxPool1	2×2	2	76×76×128
CSP2	3×3/1×1	1	76×76×128
MaxPool2	2×2	2	38×38×256
CSP3	3×3/1×1	1	38×38×256
MaxPool3	2×2	2	19×19×512

### III. IMPROVED NETWORK MODEL OF YOLOV4-TINY

#### A. Improved K-means Clustering Algorithm

In order to improve the recall rate of the algorithm model, the anchor box mechanism is introduced [13]. The anchor box is an initial candidate box with a fixed size and aspect ratio. The design of the anchor box will directly affect the convergence difficulty of the loss function in model training, thus affecting the detection accuracy and speed of the model.

The size and aspect ratio of anchor boxes are affected by the size of all real boxes in the data set. Therefore, for different data sets, it is necessary to select the appropriate anchor boxes in order to make the model training stable and accelerate the convergence.

In the process of clustering iteration, the K-means clustering algorithm [26] uses distance as the similarity index to find K classes in the given dataset, and the center of each class is obtained according to the mean value of all data points in the class. The standard K-means uses Euclidean distance to calculate the distance between two samples. The distance of the original YOLOv4-Tiny algorithm is based on the IOU when clustering and selecting the candidate box on the MS COCO data set [27]. In this section, GIoU is used as the distance basis because the IOU can't distinguish the different alignment between two objects.

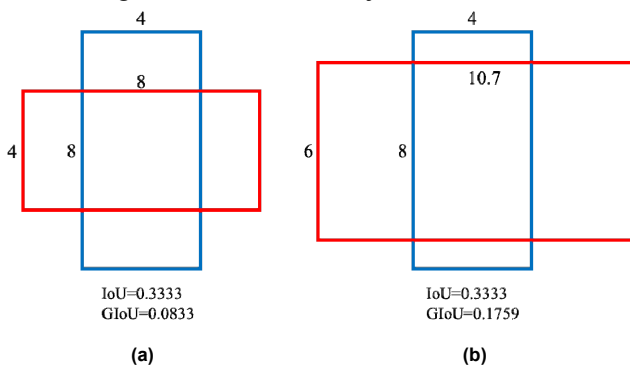


FIGURE 2. Schematic diagram of the difference between GIoU and IoU

As shown in Figure 2, the aspect ratios of the two red boxes are different. The aspect ratio of the left red box is 2 and the aspect ratio of the right red box is 1.78, but the IOU of the red box and the blue box in Fig. 2(a) and 2(b) are both 0.333. Therefore, only using IOU can not distinguish the difference in the aspect ratio of the two red boxes. GIoU considers the area of the non-overlapping areas, when the two boxes do not completely intersect, GIoU introduces the non-overlapping area item, which can better reflect the

shape information. If GIoU is used as the measurement factor of distance, the GIoU values in Figures 2(a) and 2(b) are 0.0833 and 0.1759 respectively.

The calculation formula is shown in equation (1).

$$D(x_i) = 1 - GIoU[(w_i, h_i), (w_c, h_c)] \quad (1)$$

Where  $D(x_i)$  is the distance between the width and height  $(w_i, h_i)$  of each sample point  $x_i$  and the width and height  $(w_c, h_c)$  of the clustering center.

$$GIoU = IoU - \frac{|C \setminus (A \cup B)|}{|C|} \quad (2)$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

$IoU$  is the intersection ratio of sample box size and cluster box size,  $A$  represents the size of the sample box,  $B$  denotes the size of the cluster box,  $C$  is the minimum size of the rectangular box that simultaneously surrounds the two boxes  $A$  and  $B$ , and  $C \setminus (A \cup B)$  represents the difference between the area of the rectangular box  $C$  and the union area of the two boxes  $A$  and  $B$ .

Based on TT1000K [28] traffic sign data set, this paper uses K-Mean-GIoU algorithm to do anchor box clustering on traffic sign data set, and gets six anchor boxes, which are (5,5), (7,8), (9,10), (12,14), (18,20), (29,31).

#### B. Large Scale Feature Map Optimization Strategy

The output layer of the original YOLOv4-Tiny network is a two-scale feature map with 32 times and 16 times down sampling. Generally speaking, receptive field [29] refers to the area acting on the input image, so the deeper the network layer is, the larger the receptive field is. The receptive field of deep feature map with low resolution is larger, which is used to detect large targets. The receptive field of shallow feature map with high resolution is smaller and rich in spatial information, so it is more suitable for detecting small targets. Corresponding to the proposed algorithm in this paper, the feature map with a resolution of  $19 * 19$  is responsible for detecting large targets, and the feature map with a resolution of  $38 * 38$  is responsible for detecting small targets. Since there are more small targets in the traffic sign data set, this section makes corresponding improvements.

The pooling layer MaxPool3 is deleted, the 32 times down sampling is modified to 16 times down sampling, and the original network path after network layer CSP3 is re-routed to network layer CSP2. After 8 times down sampling, the two scale output feature graph sizes of YOLOv4-Tiny detection network are changed from the original  $19 * 19$  and  $38 * 38$  to  $38 * 38$  and  $76 * 76$ , so as to improve the detection accuracy of small targets. The improved network structure is shown in Figure 3.

**C. Improved NMS Algorithm to Screen Prediction Boxes**

The idea of non-maximum suppression algorithm [30] is that in the post-processing stage of model detection, for a target object, the one with the highest confidence score among all prediction boxes of the target is selected as the benchmark prediction box, and a threshold is set. For the prediction box that overlaps with the benchmark prediction box, the prediction box whose overlap degree is greater than the threshold value is deleted, the prediction box whose overlap degree is less than the threshold value is retained, and all the prediction boxes without overlap are retained.

NMS algorithm has obvious disadvantages. Firstly, it needs to set a threshold manually, which is determined by experience. Secondly, when similar targets are dense and the detected objects are highly overlapped, the overlap degree between prediction boxes is high. NMS algorithm is easy to delete the prediction box belonging to another target, resulting in missed detection. As shown in Figure 4, the two targets of speed limit of 20 km/h of traffic signs and prohibition of trucks are highly overlapped, and only one target box is retained after NMS algorithm processing, which leads to missed detection.

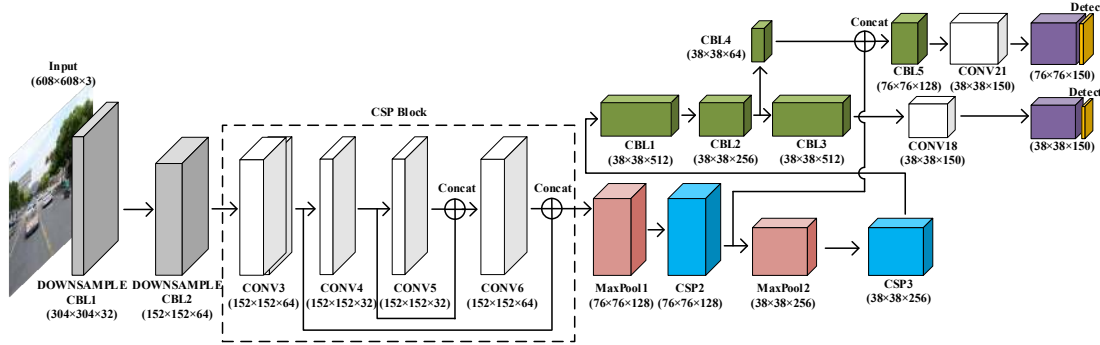


FIGURE 3. Traffic sign detection network structure diagram based on YOLOv4-Tiny improved large-scale feature map



FIGURE 4. Schematic diagram of missed targets in NMS algorithm detection

Aiming at the existing problems of NMS algorithm, this paper proposes an improved algorithm based on soft-NMS [25] to complete the screening task of YOLOv4-Tiny prediction box. In the execution process of the original soft-NMS algorithm, the confidence score of the prediction box larger than the threshold value was reduced by the score reset function, rather than deleted directly. Based on this method, when the unprocessed prediction box overlaps most of the benchmark prediction box, the box will have a low confidence score. On the contrary, if there is only a small amount of overlap, the original confidence score will not be significantly affected.

The fractional reset function can be expressed in two forms. One is linear weighting, as shown in equation (4), which can be expressed as  $f_x(IoU(M, b_i))$ . The other is Gaussian weighting, as shown in equation (5), which is expressed as  $f_g(IoU(M, b_i))$ .

$$f_x(IoU(M, b_i)) = \begin{cases} 1 & IoU(M, b_i) \leq N_i \\ 1 - IoU(M, b_i) & IoU(M, b_i) \geq N_i \end{cases} \quad (4)$$

$$f_g(IoU(M, b_i)) = e^{-\frac{IoU(M, b_i)^2}{\sigma}}, \forall b_i \notin D \quad (5)$$

In the experiments, Gaussian weighting is mostly used. Based on the Gaussian weighting function, a new fractional reset function is proposed in this paper, as shown in equation (6), which can be expressed as  $f_i(IoU(M, b_i))$ .

$$f_i(IoU(M, b_i)) = e^{-A \cdot \arctan\left(\frac{IoU(M, b_i)^2}{\sigma}\right)}, \forall b_i \notin D \quad (6)$$

Where A is an artificially set coefficient. According to the conclusion of experiments conducted by the author of Soft-NMS, when  $\sigma = 0.5$ , Soft-NMS has better performance.

In order to ensure that when the overlap degree  $IoU(M, b_i)$  between the prediction box  $b_i$  and the benchmark prediction box  $M$  is small, its original confidence score will not be greatly affected, the experiment shows that the effect is better when  $A=2$ . At this time, the image comparison of the exponential part function of the score reset function is shown in Figure 5.

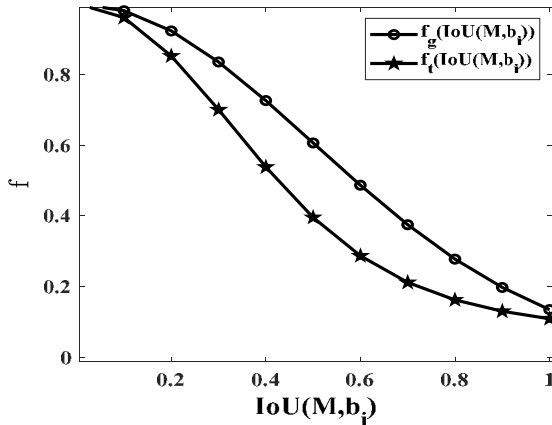


FIGURE 5. Image comparison of score reset function

As can be seen from Figure 5, when  $IoU(M, b_i)$  is small, the attenuation range of y value is small, When  $IoU(M, b_i)$  is large, the decrease range of  $f_l(IoU(M, b_i))$  is larger than that of  $f_g(IoU(M, b_i))$ , that is, the attenuation range of fractional reset function is larger, which indicates that when

the overlap  $IoU(M, b_i)$  between prediction box  $b_i$  and benchmark prediction box  $M$  is larger, the confidence fractional attenuation of prediction box  $b_i$  is more serious. For prediction box with high overlap, it is beneficial for the confidence score of the prediction box  $b_i$  to decay below the score threshold, so as to speed up the screening process of prediction box.

#### IV. MODEL PERFORMANCE EVALUATION

##### A. Data Set and Laboratory Environment

There are 9176 pictures in TT100K data set in China, with 221 kinds of annotation categories. The resolution of the images is  $2048 \times 2048$ . Because of the high resolution of the original image, the original image is cropped in this experiment, and the scale of the cropped image is  $608 \times 608$ . Due to the serious imbalance of data amount among various categories in the data set, only 45 categories of traffic signs with a large amount of data are selected for recognition in this experiment, and the training set and test set are divided according to the ratio of 8:2, in which the number of pictures in the training set is 7262, and the number of pictures in the test set is 1914. Figure 6 shows the traffic sign category in the TT100K data set, in which "\*" is used to represent other numbers of signs of the same type. For example, the speed limit sign "pl\*" includes pl25, pl30, pl35, etc.

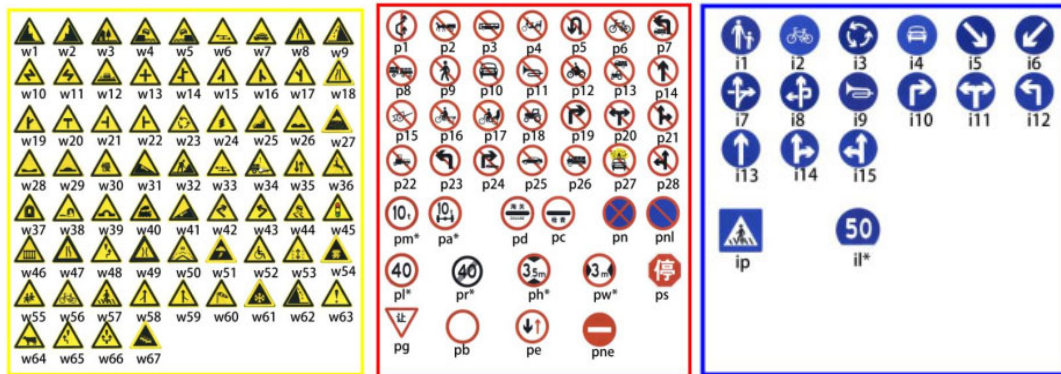


Figure 6. Traffic sign categories

The experimental training environment configuration is as follows: Intel Xeon W-2135 CPU, NVIDIA GeForce RTX 2080TI graphics card, 32G memory, Windows10 operating system, Darknet deep learning framework, CUDA Toolkit version is 10.2.

##### B. Experimental Scheme and Result Analysis

This paper focuses on the realization of the light weight traffic sign recognition algorithm model, which provides

the possibility for practical application. Taking YOLOV4-Tiny as the benchmark, different innovative strategies are combined to perform training and performance statistics on the TT100K data set, so as to improve the detection accuracy of the model as much as possible under the premise of ensuring real-time performance.

The test results of different algorithm models based on TT100K data set are shown in Table II.

TABLE II  
TEST RESULTS AFTER DIFFERENT IMPROVEMENT STRATEGY COMBINATIONS

improvement strategy	Different YOLOv4-Tiny algorithm models					
	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5	
Improved clustering Anchor		√	√	√	√	
Large-scale optimization strategy			√	√	√	
Original Soft-NMS				√		
Improved Soft-NMS					√	
	mAP/%	46.34	47.16	51.02	51.39	<b>52.07</b>
evaluation indicator	Recall/%	57.23	59.61	62.01	62.32	<b>64.52</b>
	FPS/ f/s	87.9	88.2	87.6	86.4	<b>87.1</b>

As shown in Table II, on the TT100K data set, the mean average precision (mAP) value of the original YOLOv4-Tiny (experiment 1) was 46.34%, and the mAP value of the cluster anchor box (experiment 2) was 47.16%. The improvement effect was obvious, and the recall rate was increased by 2.38 percentage points. Because there are more small-scale targets in the traffic sign data set, the combination of the large-scale optimization feature map strategy and the improved clustering anchor strategy proposed in this paper achieves good detection effect, as shown in experiment 3, the mAP value reaches 51.02%, which is 4.68% higher than that in experiment 1, and the recall rate reaches 62.01%, which is 4.78% higher than that of experiment 1. This is because the proportion of small targets in the traffic sign data set is higher. After introducing this optimization strategy, the shallow feature semantic information extracted from the feature extraction, especially the learning and training of small targets is more in-depth, so as to improve the detection performance of the algorithm model.

Considering the partial occlusion or overlap of traffic sign targets in the data set, soft-NMS algorithm is introduced in experiment 4, and the mAP value reaches 51.39%. Compared with experiment 3, the accuracy of the model is further improved. It can be seen that soft-NMS algorithm has a positive impact on the accuracy of model improvement, so on this basis, experiment 5 adopts the improved soft-NMS strategy, the mAP value reaches 52.07%, 0.68% higher than experiment 4, 5.73% higher than the original YOLOv4-Tiny (experiment 1), the recall rate reaches 64.52%, 2.2% higher than Experiment 4, 7.29% higher than experiment 1, and the improvement range is obvious. Moreover, the FPS value of the algorithm models is maintained at about 87f/s, which is basically unaffected.

For comparison, the test results of experiment 1 and experiment 5 on the TT100K data set are counted, as shown in Table III, where class represents the categories in the data set, AP1 represents the average precision (AP) value of each category in experiment 1, and AP2 represents the AP value of each category in experiment 5.

TABLE III  
STATISTICS OF AP VALUE OF 45 CATEGORIES IN TT100K DATASET

class	AP1\%	AP2\%	class	AP1\%	AP2\%
i2	67.08	<b>69.27</b>	pl120	42.43	<b>51.06</b>
i4	84.88	82.37	pl20	28.63	26.00
i5	89.32	88.93	pl30	31.13	<b>33.25</b>
il100	40.37	<b>72.45</b>	pl40	41.63	<b>46.27</b>
il60	68.89	<b>78.64</b>	pl5	59.27	<b>63.24</b>
il80	44.47	<b>60.13</b>	pl50	36.54	<b>41.56</b>
io	62.79	<b>78.06</b>	pl60	40.07	<b>41.77</b>
ip	82.12	80.27	pl70	40.31	33.04
p10	48.67	<b>52.47</b>	pl80	41.67	<b>47.45</b>
p11	34.79	<b>68.63</b>	pm20	30.76	<b>36.81</b>
p12	25.76	<b>36.32</b>	pm30	17.08	13.44
p19	25.68	14.66	pm55	51.69	33.92
p23	53.21	44.30	pn	52.72	<b>83.90</b>
p26	47.06	<b>63.03</b>	pne	89.94	<b>90.01</b>
p27	33.18	<b>46.01</b>	po	36.93	<b>55.62</b>
p3	35.02	<b>61.80</b>	pr40	73.63	<b>79.75</b>
p5	73.72	72.73	w13	21.05	<b>28.81</b>
p6	9.71	<b>11.93</b>	w32	28.40	22.82
pg	87.12	66.11	w55	34.98	29.65
ph4	21.66	<b>23.39</b>	w57	54.39	<b>60.08</b>
ph4.5	59.73	<b>68.04</b>	w59	54.10	<b>64.74</b>
ph5	19.04	19.01	wo	32.69	24.16
pl100	57.10	<b>66.36</b>			

According to the analysis of table III, the model test results of experiment 5 using the improved strategy in this paper perform better than the model test results of experiment 1 using the original YOLOv4-Tiny. The average precision of more than 65% of the 45 categories (black/red and bold representation) in TT100K dataset has been improved to varying degrees, and the AP values of 8 categories (red and bold representation) are improved by more than 15%, which are il100, il80, io, p11, p26, p3, pn

and po. In experiment 1 and experiment 5, the PR curves of these 8 types are shown in Figure 7.

In Figure 7, the red curve marked with circle symbol is the PR curve of experiment 1 (the original YOLOv4-Tiny algorithm model) on each category, while the black curve marked with triangle symbol is the PR curve of experiment 5 (the improved YOLOv4-Tiny algorithm model) on each category after the combination of optimization strategies proposed in this paper. The area under the PR curve and

enclosed by the coordinate axis of each traffic sign category is the AP value of the corresponding category. The higher the curve is to the upper right, the larger the area enclosed by the curve and the coordinate axis is, the higher the corresponding AP value is and the better the performance is. As can be seen from the figure, the algorithm model of experiment 5 has achieved good detection performance in all the 8 traffic sign categories, and the AP value of each category has been significantly improved compared with experiment 1.

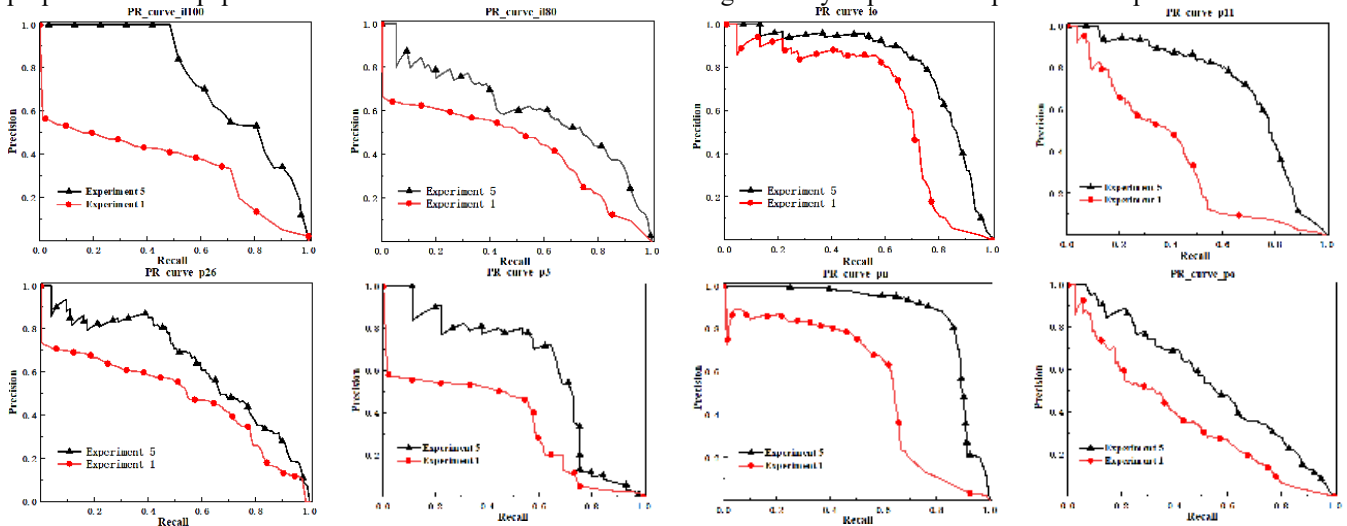


Figure 7. Comparison of PR curves of different categories

TABLE IV  
COMPARISON OF OVERALL DETECTION PERFORMANCE

algorithm	mAP%/(IoU=0.5)	FPS/s	Model size/Mb
YOLOv3-Tiny	44.15	60.4	34.3
YOLOv4-Tiny	46.34	87.9	23.4
Improved YOLOv4-Tiny	52.07	87.1	24.7

TABLE V  
OTHER EVALUATION INDICATORS

algorithm	TP	FP	TN	Recall	Accuracy	F1-Measure
YOLOv3-Tiny	2578	1460	2543	50.34%	39.17%	56.29%
YOLOv4-Tiny	2931	1785	2190	57.23%	42.44%	59.65%
Improved YOLOv4-Tiny	3304	1405	1817	64.52%	50.63%	67.22%

The overall detection performance and some other evaluation indicators of the improved YOLOv4-Tiny algorithm, YOLOv3-Tiny and YOLOv4-Tiny are compared. The results are shown in table IV and table V.

As can be seen from table IV, in terms of detection accuracy, the improved YOLOv4-Tiny algorithm improves 7.92% and 5.73% respectively compared with YOLOv3-Tiny algorithm and YOLOv4-Tiny algorithm. In terms of detection speed and model size, it has obvious advantages over YOLOv3-Tiny.

It can be seen from table V that some evaluation indicators of the improved YOLOv4-Tiny algorithm are obviously better than the other two algorithm models. Among them, F1-measure is the weighted harmonic average of precision

and recall. The two indicators of precision and recall are integrated to measure the quality of the algorithm model. The higher the F-measure value, the more effective the algorithm model will be.

Figures. 8(a1), 8(b1) and 8(c1) show the detection effect of the YOLOv4-Tiny algorithm model in the real scene, and 8(a2), 8(b2) and 8(c2) show the detection effect of the improved YOLOv4-Tiny algorithm model in the real scene. The detection results show that compared with the original YOLOv4-Tiny algorithm, the proposed method has better target location and recognition effect and higher detection accuracy. Moreover, the selected test images are real road scenes, which also proves the high robustness of this method.



Figure 8. Detection renderings of the improved algorithm model in different scenes

## V. CONCLUSION

In this paper, based on the framework of YOLOv4-Tiny algorithm, aiming at the characteristics of traffic sign data set and the shortcomings of original YOLOv4-Tiny algorithm in traffic sign detection, three feasible improvement strategies are proposed: improved K-means clustering algorithm to generate anchor box suitable for traffic sign data set, large-scale optimization feature map strategy, and improved soft-NMS algorithm to filter prediction box aiming at the shortcomings of NMS algorithm in the post-processing stage of the model, so as to improve the detection accuracy on the premise of ensuring the real-time traffic sign recognition. The experiment shows that the mAP value and recall value of the improved YOLOv4-Tiny algorithm in TT100K data set are 5.73% and 7.29% higher than the original YOLOv4-Tiny algorithm model respectively, reaching 52.07% and 64.52%, which greatly improves the detection accuracy and recall rate, and provides a certain experimental basis for the practical application of the subsequent traffic sign recognition algorithm.

## REFERENCES

- [1] Y. Jin, Y. Fu, W. Wang, J. Guo, C. Ren and X. Xiang, "Multi-feature fusion and enhancement single shot detector for traffic sign recognition," *IEEE Access*, vol. 8, pp. 38931-38940, 2020.
- [2] N. Tang, H. Liu, K. Yue, W. Li and X. Yue, "Automatic classification for corneal ulcer using a modified VGG network," presented at the *2020 Int. Conf. Artificial Intelligence and Comput. Eng.*, Beijing, China, Oct. 23-25, 2020.
- [3] R. Tanadi, E. Husni and A. R. Emir, "Camera latency review and parameters testing for real-time object detection implementation," presented at the *6th Int. Conf. Interactive Digital Media*, Bandung, Indonesia, Dec. 14-15, 2020.
- [4] S. Acharya, P. K. Nanda, "Adjacent LBP and LTP based background modeling with mixed-mode learning for foreground detection," *Pattern Anal. Appl.*, pp. 1-28, Feb. 2021.
- [5] F. Shao, X. Wang, F. Meng, T. Rui, et al, "Real-time traffic sign detection and recognition method based on simplified Gabor wavelets and CNNs," *Sensors (Basel)*, vol.18, no. 10, p. 3192, 2018.
- [6] A. Salhi et al. "Traffic signs recognition using HP and HOG descriptors combined to MLP and SVM Classifiers," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 11, 2017.
- [7] C. Cortes, V. Vapnik, "Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273-297, Sep. 1995.
- [8] Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119-139, Aug. 1997.
- [9] D. Cireşan, U. Meier, J. Masci and J. Schmidhuber, "A committee of neural networks for traffic sign classification," *Int. Joint Conf. on Neural Networks*, San Jose, CA, USA, Jul. 31- Aug. 5, 2011.
- [10] D. Cireşan, U. Meier, J. Masci and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*,



- vol. 32, pp. 333-338, 2012.
- [11] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, Jun. 23-28, 2014.
- [12] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448, 2015.
- [13] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE T. Pattern Anal.*, vol. 39, no. 6, pp. 1137-1149, Jun. 2017.
- [14] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: unified, real-time object detection," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 27-30, 2016.
- [15] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 21-26, 2017.
- [16] A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," *arXiv preprint arXiv: 2004.10934*, 2020.
- [17] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," *arXiv preprint arXiv: 1804.02767*, 2018.
- [18] W. Liu, D. Anguelov, D. Erhan, *et al.*, "Ssd: single shot multibox detector," *European Conference on Computer Vision (ECCV)*, Springer, Cham, DE, vol. 9905, pp. 21-37, Oct. 2016.
- [19] Z. R. Zuo, K. Yu, Q. Zhou, *et al.*, "Traffic signs detection based on faster R-CNN," *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, Atlanta, GA, USA, Jun. pp. 286-288, 2017.
- [20] S. You, Q. Bi, Y. M. S. D. Liu, *et al.*, "Traffic sign detection method based on improved ssd," *Information*, vol. 11, no. 10, pp. 475, Oct. 2020.
- [21] T. T. Yang, X. Long, A. K. Sangaiah, *et al.*, "Deep detection network for real-life traffic sign in vehicular networks," *Computer Networks*, vol. 136, no. 8, pp. 95-104, 2018.
- [22] Y. Yuan, Z. T. Xiong and Q. Wang, "VSSA-NET: vertical spatial sequence attention network for traffic sign detection," *IEEE transactions on image processing*, vol. 28, no. 7, pp. 3423-3434, Feb. 2019.
- [23] H. B. Zhang, L. F. Qin, J. Li, *et al.*, "Real-time detection method for small traffic signs based on Yolov3," *IEEE Access*, vol. 8, pp. 64145-64156, Mar. 2020.
- [24] H. Rezaatfighi, N. Tsoi, J. Y. Gwak, *et al.*, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658-666, June, 2019.
- [25] N. Bodla, B. Singh, R. Chellappa, *et al.*, "Soft-NMS - improving object detection with one line of code," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5561-5569, 2017.
- [26] L. J. Tao, L. Y. Hong and H. Yan, "The improvement and application of a K-means clustering algorithm," *IEEE Int. Conf. on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, China, Jul. 5-7, 2016.
- [27] T. Y. Lin *et al.*, "Microsoft coco: Common objects in context," *European conference on computer vision (ECCV)*, Springer, Cham, pp. 740-755, Sep. 2014.
- [28] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li and S. Hu, "Traffic-sign detection and classification in the wild," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 27-30, 2016.
- [29] Y. G. Liu, J. Z. Yu and Y. H. Han, "Understanding the effective receptive field in semantic image segmentation," *Multimed. Tools Appl.*, vol. 77, no. 17, pp. 22159-22171, 2018.
- [30] L. Cai, B. Zhao, Z. Wang, J. Lin, C. S. Foo, M. S. Aly, V. Chandrasekhar, "MaxpoolNMS: getting rid of NMS bottlenecks in two-stage object detectors," *2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, Jun. 15-20, 2019.



**LANMEI WANG** was a professor at school of physics and optoelectronic engineering in Xidian University, China. She received the Master of Science degree in radio physics in 2002 and the Ph.D. degree in information engineering in 2005, both from Xidian University, Xi'an Shaanxi, China. Her research interests include in polarization sensitive array signal processing and the application of neural network and artificial intelligence in target detection and recognition.



**KUN ZHOU** was born in Shaanxi Province, China, in 1996. He received the B.S. degree from the Shaanxi University of Technology, Shaanxi, China, in 2018. He is currently pursuing the M.S. degree with the School of Physics and Optoelectronic Engineering, Xidian University, Shaanxi, China. His current research interest is the application of neural network in deep learning and target detection.



**ANLIANG CHU** was born in Shandong Province, China, in 1997. He received the B.S. degree from the China University of Petroleum (East China), China, in 2018. He is currently pursuing the M.S. degree with the School of Physics and Optoelectronic Engineering, Xidian University, Shaanxi. His current research interest is the application of neural network in traffic sign recognition.



**GUIBAO WANG** was a professor at School of Physics and Telecommunication Engineering, Shaanxi University of Technology, China. He received the Master of Science degree in computational mathematics in 2003 and the Ph.D. degree in information engineering in 2014, both from Xidian University, Xi'an Shaanxi, China. His research interests focus on the electromagnetic vector sensor array signal processing and the intelligent perception and target recognition.



**LIZHE WANG** was born in Shaanxi Province, China, in 1993. He received the B.S. degree in Shaanxi University of Technology, Shaanxi, China, in 2017, and M.S. degree in Kunming University of Science and Technology, Yunnan, Kunming, in 2010. Currently pursuing the Ph.D. degree in radio physics at Xidian University, Xian, Shaanxi, China. His current research interest is deep learning and image processing.