Research on Traffic Sign Object Detection Algorithm Based on Deep Learning

Mingyang Sun, Ying Tian

Abstract—**Traffic mark detection and identification play a key character in the development of driverless and intelligent transportation systems, offering significant assistance in ensuring the safety of people's daily travels. However, the detection effect of traffic signs is affected by many target categories, small targets, and low recognition accuracy, making traffic sign detection more challenging than target detection in general scenarios. In this paper, an improved YOLOv7 network (YOLOv7-COORD) is entered. Foremost, increase CBAM attention module at the connection between backbone and neck network of YOLOv7 to enhance the expression ability of neural networks through the attention mechanism, emphasizing important features and ignoring minor features to enhance the efficiency and precision of the network. Secondly, By adding CoordConv before the upsampling of the neck and before the detection head output, the network can better feel the location message in the characteristic map. Finally, a detection head generated by the low-level, high-resolution characteristic map is added to enhance the recognition accuracy of small target object. The abundance of experimental data demonstrates that the impression of the improved YOLOv7-COORD model is superior to that of the original YOLOv7 model, and the average accuracy of (mAP@0.5) on TT100K datasets is 3.2% higher than that of YOLOv7, reaching 85.4%. In summary, the improved YOLOv7-COORD model can better detect targets in traffic sign images.**

Index Terms—**Traffic sign detection, YOLOv7, CBAM, CoordConv**

I. Introduction

In recent years, in order to solve various social problems often caused by traffic safety and traffic jams, research on driver assistance systems [1] and automatic driving has been favoured by many researchers. The recognition of traffic signs is the most essential part of automatic transmission. [2] Traffic signs contain rich semantic information, which reminds or warns drivers on the road to guarantee the security of people and drivers at all times. Therefore, only rapid, efficient, and accurate identification of traffic lights, traffic signs, and other essential information can improve the safety of automatic driving.

The traditional traffic mark goal detection algorithm primarily focuses on feature extraction and classification. It involves segmenting the color space based on the shape and edges of the traffic sign, combined with a feature extraction method to extract relevant characteristics. Then, the classifier completes the feature classification to realize the recognition of the traffic sign. Wang Bin [3] et al.combined three feature extraction methods, the LBP feature, HOG feature, and GIST feature, carried out dimension reduction of data through primary component analysis (PCA) and then used support vector machine to complete target training and recognition, which improved the classification accuracy. Still, the accuracy of category recognition of some traffic marks was low. Wang Yan [4] et al.extracted the Zernike moment invariant feature of the image and then completed the recognition of traffic signs by support vector machine (SVM), which strengthened the recognition rate of traffic signs in a sophisticated environment. However, the initial workload was large, and the use was more complicated. Liang Minjian [5] et al. used the HOG-Gabor characteristic fusion method to obtain feature vectors and then realized target recognition through a classifier, which enhanced the correct recognition ratio of targets, but the robustness was low. The traditional traffic sign detection algorithm, despite its increasing accuracy, still suffers from drawbacks such as high computational complexity and intricate operations.

The continuous advancement of computing power has propelled deep learning-based object detection and measurement methods [6] to the forefront as the preferred choice among researchers. The researchers have put forward a one-stage goal detection algorithm and a two-stage goal detection algorithm. Girshick et al. [7] proposed R-CNN, regarded as the forerunner of a two-stage algorithm. He proposed to extract candidate regions first and then using convolution neural network and SVM classifier for classification prediction. On this basis, researchers have successively proposed more efficient two-stage detection algorithms such as Fast R-CNN [8], Faster R-CNN [9], Mask R-CNN [10], etc. However, due to its characteristics, the two-stage algorithm cannot well satisfying the demands of real-time detection. The YOLO (You Only Look Once) series, proposed by Redmon et al. [11], introduces an innovative approach to target detection. The one-stage object detection model fuses the steps of candidate region extraction into feature extraction, and can complete the position regression and classification tasks only by picking up the features of all pixels in the input image. The one-stage prediction reasoning process is relatively simple, which realizes the purpose of speeding up but also increases the difficulty of detection and reduces the detection accuracy. It perceives detection as a restoration problem and replaces traditional region selection with the utilization of a preset anchor frame and the square formula. After this, YOLO algorithms such as YOLOv7,

Manuscript received December 1, 2023; revised June 7, 2024.

Mingyang Sun is a postgraduate student majoring in software engineering at School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan Liaoning 114051, China. (e-mail:qq940518909@163.com)

Ying Tian is a professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Liaoning 114051, China. (corresponding author to provide phone: +8613898015263; e-mail:astianying@126.com)

YOLOv8, and SSD [12] detection algorithms were proposed successively, and their demonstration in terms of precision and speed became better and better.

Scholars at home and abroad have explored different research schemes for the detection of traffic marks. Sudha et al. [13] use a new method of stochastic gradient dynamic sequence and shape feature extraction for detection and then use convolution neural networks (convolution neural networks, convolution neural networks, etc). The classifier classifies the training output tags and finally converts traffic signs into audio signals during the training phase and testing phase to help the visually impaired solve problems. Han et al. [14] proposed a Faster R-CNN for small-target traffic mark detection. First, a small range pointer network (RPN) extracts the suggestion box. Then, Faster R-CNN is combined with an online complicated example mining (OHEM) algorithm to enhance the capacity to locate small targets. Wan et al. [15] used the improved YOLOv3 for model pruning and multi-scale prediction, which improved the detection velocity while the quantity of algorithm parameters is reduced. Still, the real-time detection timeliness was not high. Yin Jinghan et al. [16] Used improved YOLOv5 to identify traffic signs in haze scenarios, reduce the depth of the feature pyramid, limit the maximum under-sampling multiple, adjust the depth of feature transmission of the residual module, and other methods to enhance the detection precision.

To sum up, because of the high real-time requirements of traffic sign detection, the current model cannot meet the demand of quick detection velocity and accurate detection precision. Therefore, this paper uses YOLOv7 model as the benchmark model. Considering that traffic signs have a relatively small image count in the image, this paper adds a small goal detection layer to enhance the sensitivity to small goals. Meanwhile, the Convolution Block Attention Module (CBAM) is integrated into the model. By learning the attention weight value from the characteristic map along the space and channel, the attained weight is multiplied with the original characteristic map to enhance the network's attention to the vital characteristic information in the original characteristic and thus enhance the perception ability of the neural network. The convolution in the neck layer is changed into CoordConv convolution. Compared with traditional convolution, CoordConv adds coordinates to the convolution so that the convolution can perceive spatial information, thus improving the model's capability to identify traffic marks. Finally, experiments are carried out on the public TT100K datasets to demonstrate the effectiveness of the enhanced YOLOv7-COORD algorithm.

II. Related algorithms

YOLOv7 [17] model is a single-stage goal detection network entered in July 2022. It currently stands as one of the most advanced goal detection models, surpassing the majority of known counterparts both velocity and precision within the range of 5 ~ 160fps. The network improves algorithms and uses more efficient models for faster processing speeds. Compared with the previous YOLOv4 [18] and YOLOv5 [19], the YOLOv7 model has better detection speed and accuracy and is more flexible and easy to use. At the same time, it can process pictures more quickly and is more suitable for real-time monitoring. Therefore, this model is very appropriate for the deployment of in-vehicle equipment with high real-time requirements, small memory, and low computing resources for traffic sign detection. The model of YOLOv7 is mainly separated into four parts: (1) Input: After the input image is scaled, it meets the input size requirements of Backbone. (2)Backbone: combine images of different scales in depth to form a series of feature maps; (3)Neck: multi-scale image features are fused, and then the combined feature images are input into the prediction layer for prediction; (4)Head: The input image characteristics are predicted, and then the corresponding bounding box is output, as well as the prediction category and confidence score. For the input image of any size, the model first preprocesses the data and adaptive scales the size of the input image to 640×640×3 by using rectangular reasoning. YOLOv7 uses a more complex network structure based on YOLOv5, such as DarkNet-67, which is more profound than the DarkNet-53 used by YOLOv5 and can improve the model's performance. The E-ELAN [17] structure in the characteristic extraction and characteristic blend module enhances the network's learning ability.

Although the YOLOv7 algorithm framework performs well in standard target detection fields, its application to traffic sign target detection in complex road environments still faces many difficulties and limitations. The MPConv structure of YOLOv7 can only be used for traditional pooling operations. However, the sampling position around each position cannot be adjusted adaptively during pooling, so the robustness of the target deformation and position change is poor. In terms of loss function, the size of the traffic sign is minimal. For the whole image, the pixel is small, resulting in the deviation of the position and shape of the object, which has a significant impact on the default coordinate loss. On the basis of the classification loss function, cross-entropy loss, regardless of sample quality, treats all samples equally because of the different distances and illumination degrees of traffic signs in the datasets and has a particular impact on detection accuracy. This paper makes improvements based on YOLOv7 to solve the problems above.

III. Improvements

A. CBAM

The common CA [20] and SE [21] attention mechanisms are both attention mechanisms that calculate the weight of the channel dimension. At the same time, CBAM [22] is a comprehensive attention mechanism that consist of channel and space [23] and can serialize the information of attention characteristic maps in both channel and space. The combination of maximum pooling and average pooling is used to enrich the feature message. The traffic marks that need to be detected in the traffi mark datas take up a small ratio of the picture, which is a tiny target Contributing the CBAM attention mechanism into my network model can improve the weight of the traffic mark characteristic region and reduce the weight of the background feature when the network performs feature learning. The structure of CBAM's attention mechanism is shown in Fig. 1 CBAM consists of two attention modules. Channel attention module: The characteristic map is compressed by global pooling operations, and the attention weight is calculated by

multi-layer perceptron to focus on crucial information, channel attention, spatial attention modules: The convolution operation is compressed, and the activation function is weighted and summed, focusing on spatial position information. The experimental data demonstrate that the speed and accuracy of this method outperform other methods for detecting traffic sign targets.

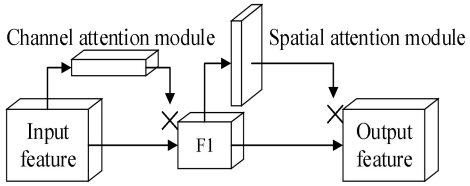

Fig. 1. The structure of CBAM

Channel attention calculation formula:

$$M_c(F)=\delta(MLP(avgpool(F)+MLP(maxpool(F))) \quad (1)$$

Spatial attention calculation formula:

$$M_s(F)=\delta\{f[avgpool(F_1);(maxpool(F_1))] \quad (2)$$

In the above formula, MLP is the sigmoid function, avgpool is average pooling, maxpool is maximum pooling ";" is concat, and f is convolution.

B. The small target detection layer

Since the target pixels to be tested in the traffic sign data set are small, YOLOv7 has a prominent ability to detect large targets. If the original YOLOv7 detection model continues to be used, it may cause the loss of small goal information after multiple down-sampling. Therefore, in an effort to enhance the detection precision of small goals and effectively increase the characteristic extraction capability of small goals, In this paper, based on not changing the scale of other characteristic maps, a P2 small goal detection stratum with a resolution of 160×160 is increased to the neck network to strengthen the detection precision of small traffic marks by YOLOv7. The position of the added small goal detection layer is located after the original last feature fusion module, as shown by the dashed red line in Fig. 4. In Fig. 4, the detection layers P2, P3, P4, and P5 correspond to 4-fold, 8-fold, 16-fold, and 32-fold down-sampling characteristic maps, respectively. The 4-fold down-sampling characteristic map has a small receptor field. It includes more shallow semantic characteristics of traffic signs, which can retain small goal characteristics to the greatest extent, thus improving goal detection precision. In addition, because of the abundant number of small goals in the road identification data set and the relatively small width and height sizes, the anchor of the original YOLOv7 model cannot meet our needs, so in this paper uses K-means clustering algorithm is adopted to manufacture a set of more matching anchors, as exhibited in Table I.

TABLE I
THE ANCHOR GENERATED BY K-MEANS ALGORITHM

Feature resolution layer	Anchor size
160×160	[5,6] [8,14] [15,11]
80×80	[10,13] [16,30] [33,23]
40×40	[30,61] [62,45] [59,119]
20×20	[116,90] [156,198] [373,326]

C. CoordConv module

In an effort to give the model a better sense of location information in the characteristic map, the traditional convolution parameters are small in number, computationally efficient, and have good translation stabilization, which can better learn the essential features of the object when dealing with tasks such as detection and classification. However, traditional convolution can only be calculated through the weights and input data in the convolution kernel, and it cannot obtain the actual coordinates and position information of pixels. CoordConv [24] is a method to add spatial information in convolution neural networks. CoordConv inherits the excellent characteristics of traditional convolution and effectively solves the problem that traditional convolution has poor ability to obtain spatial information. It does this by adding an additional coordinate channel. This channel contains information about the pixels' actual position in the image, as well as their position in other feature maps. In an effort to make the convolution have the ability to perceive spatial information, two coordinate channels are added to the feature map, expressing the i and j coordinates of the primitive input, and then traditional convolution operations are carried out that the convolution can apperceive the spatial message of the characteristic map. In this way, CoordConv can better model the connections between different locations. This makes it much better at tasks that require it to deal with location awareness. This article uses CoordConv to replace the 1×1 convolution in the head layer and the REP convolution before the detection head. Traditional convolution is exhibited in Fig. 2, and CoordConv is demonstrated in Fig. 3.

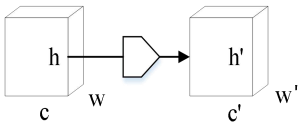

Fig. 2. Traditional convolution

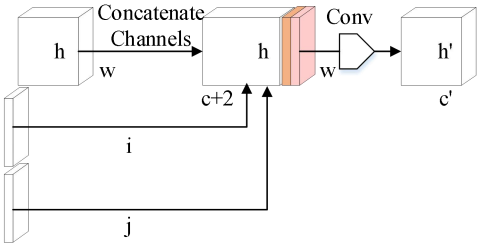

Fig. 3. CoordConv convolution

Pointing at the low detection accuracy of traffic signs in complicated road scenes, this paper takes YOLOv7 as the base model. It adds CBAM attention mechanism between the backbone and neck to highlight the characteristics of objects to be recognized. A small target detection layer with an output size of 160×160 feature map is added to the original YOLOv7 network. The map has a smaller receptive field and more abundant location communication, which increases the robustness of detection under complex road backgrounds. It is more applicable for small target detection. CoordConv replaced CBS and REPConv in the neck and head layers, respectively, to improve the detection accuracy of small targets. The improved YOLOv7-COORD model structure is displayed in Fig. 4.

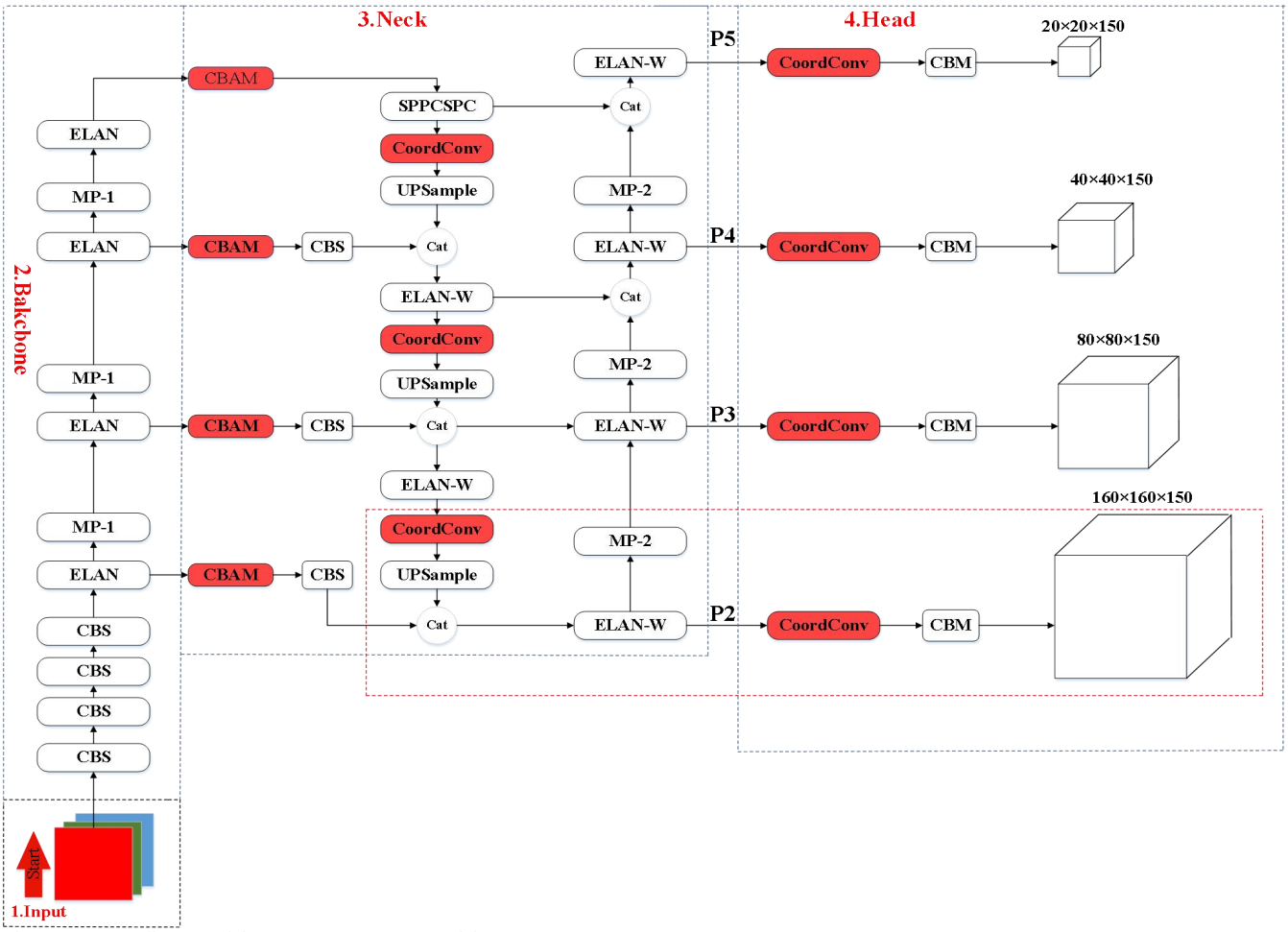

Fig. 4. Structure diagram of the YOLOv7-COORD model`

IV. Experiment and Analysis

A. Introduce to datasets

In order to better adapt to China's traffic road scene, this paper chooses to use the TT100K data set co-produced by Tsinghua University and Tencent for training [25]. The TT100K datasets have a vast foundation and rich semantic information. It includes natural street scenes captured by high-definition cameras. It can restore the first driving Angle, complex rural road sections, and various images under significant difference in strength of light and weather circumstances. It is a reliable benchmark data set for traffic sign detection. The data set contains hundreds of traffic sign categories, but most signs have few instances. Therefore, this paper first screens the data set and obtains 45 categories with more than 100 instances for the experiment. The following are the partitioning steps.

(1) Read the json annotation file through traversal, delete the labels with less than 100 instances, and save the labels with more than or equal to 100 instances.

(2) The divided json format file is separated into three json files: training file, test file, and verification file according to 8:2:1.

(3) Convert each json file into the TXT format required for YOLOv5 training and save it.

(4) Store the marked files and pictures separately according to the division.

(5) Create a data set folder and save the newly divided data set in this folder.

After the above steps, 45 common traffic signs were obtained, in the aggregate of 9166 pictures, including 6416 training sets, 1833 test sets, and 917 validation sets, as exhibited in Fig. 5.

Fig. 5. TT100K datasets examples

B. Evaluation index

There are many kinds of labels in the multi-label picture division task, so the accuracy standard of the single label picture division task can not be used as the evaluation index. mAP was used as the evaluation index in this experiment.

$$\text{Precision}=\frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{4}$$

Precision is the scale of correctly predicted positive example data to predicted positive example data, which can influence the model's capability to identify the goal.

Recall refers to the scale of the correct positive samples to all positive samples, comparing the correctly detected samples to the actual samples. TP indicates the quantity of positive samples predicted accurately, FP indicates the quantity of positive samples mispredicted, and FN indicates the quantity of negative samples mispredicted by the model. Both accuracy and recall rate can impact the model's competence to recognize the target. Therefore, the P-R curves before and after the model modification are compared to make a more overall comparison of the expression of the network model. Fig. 6 shows the P-R curve of YOLOv7, and Fig. 7 shows the P-R curve of YOLOv7-COORD.

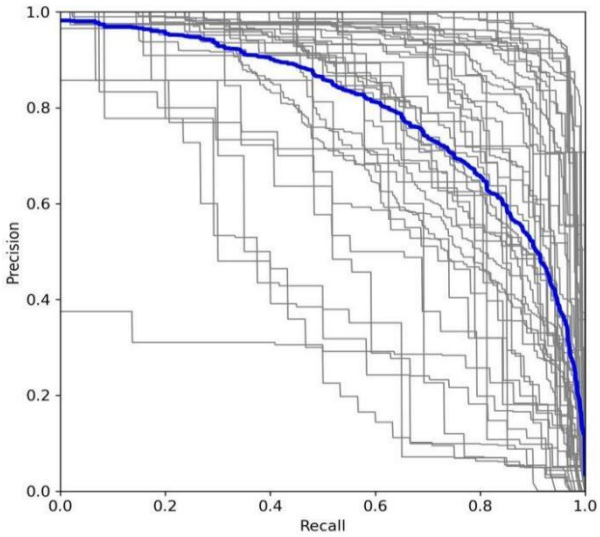

Fig. 6. YOLOv7 P-R

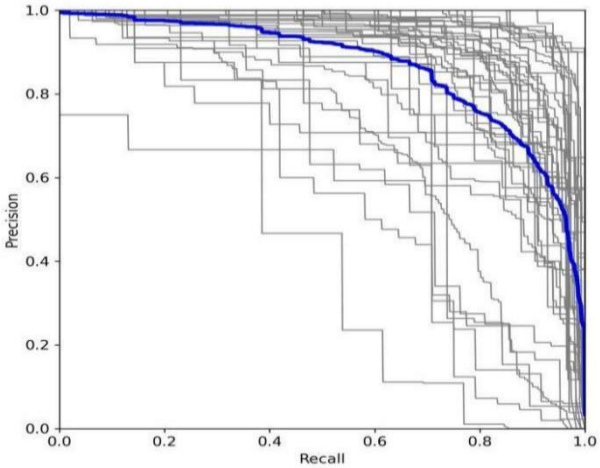

Fig. 7. YOLOv7 -COORD P-R

The P-R curve takes accuracy as the ordinate and recall rate as the abscissa, as shown in Fig. 6 and Fig. 7. The more significant the area enclosed by the blue curve, the better the model's expression. This can be obtained from Fig. 7 that the P-R curve of YOLOv7-COORD has a greater area and better detection outcome.

The precision value AP for each category is the area of the P-R curve and the coordinate axis. The mAP value is the area's average under the P-R curve for classifying these traffic signs. mAP can be used as a relatively good evaluation indicator. AP and mAP formulas are shown below:

$$AP=\frac{\sum_{i=1}^{N} P_i}{N} \tag{5}$$

$$mAP=\frac{\sum_{j=1}^{N} AP_j}{M} \tag{6}$$

In formulas (5) and (6), M represents the total amount of categories used on detection, and N represents the amount of images tested. The mAP contains mAP@0.5 and mAP@0.5:0.95, depending on what the IoU threshold you set.

C. Experimental results

Fig. 8 shows the observation of part of the detection results. In the picture ,the left side is the YOLOv7 detection diagram, and the right side is the YOLOv7-COORD detection outcome diagram. Comparing the results of the left and right pictures show that YOLOv7-COORD model has a better recognition effect and higher generalization ability for traffic sign object detection..

Fig. 8. The comparison image of YOLOv7 and YOLOv7-COORD

D. Evaluation index

All the experiments in this paper used the operating environment: 13th Gen Intel(R) Core(TM)i7-13700F processor, NVIDIA GeForce RTX 4070 Ti graphics card, 12GB graphics memory, Windows10 Professional operating system, Python3.10.0 programming language, CUDA 11.3 accelerated computing architecture, Pytorch 1.11.0 deep learning framework.

In this paper, based on YOLOv7 primitive model, the input image size is set to 640×640, the total number of iterations is set 300 epochs, the optimizer is set to SGD, the momentum parameter is set to 0.937, the learning rate is set to 0.001, and the weight attenuation coefficient is set to 0.0005.

E. Objection detection

In an effort to test the outcome of the improved YOLOv7-COORD algorithm in this paper, a comparison test is conducted between the improved algorithm and Faster R-CNN, YOLOv3, Retinanet, YOLOv5, and YOLOv7 algorithms. The recall rate and recognition accuracy of different algorithm models on the TT100K data set are shown in the table. It can be seen that Table II, the mAP of the YOLOv7-COORD model increases by 3.2% contrasted with YOLOv7. It increases by 5.2%, 6.1%, 14.5%, and 6.6% compared with YOLOv5, YOLOv3, and Faster R-CNN, which are all superior to other target detection algorithms. It is proved that the improved method put forward in this paper is effective for traffic sign target detection in complex circumstances.

TABLE II
COMPARING WITH OTHER METHODS ON TT100K

Approaches	Input size	R (%)	mAP
Fster R-CNN	1000×600	76.8	78.8%
YOLOv3	640×640	66.3	70.9%
Retinanet	1000×600	75.2	79.3%
YOLOv5	640×640	75.5	80.2%
YOLOv7	640×640	76.2	82.2%
YOLOv7-COORD	640×640	78.8	85.4%

F. Ablation experiment

In an effort to further proving the effect of the put forward improvements on the behavior of the algorithm, an ablation experiment was in progress and the consistency of input images and training parameters was maintained in all experiments. Among them, Detect head, CoordConv, and CBAM are the enhanced the ways put forward in this paper. The outcome of ablation experiments are represent in Table III. As we can seen that Table III, the first group of experiments is the original YOLOv7 model, with mAP of 82.2%. The second group of experiments added the small target detection layer, and the mAP increased by 0.8%, pointing out that added the small goal detection layer can enhance the detectable range of small goal detection. The third group of experiments added CoordConv convolution on the basis of the first group of experiments, which improved by 1.6% contrasted with the primitive YOLOv7 algorithm mAP, indicating that adding CoordConv convolution can better perceive the spatial location message in characteristic maps. The fourth set of experiments is the improved YOLOv7 algorithm proposed in this paper. On the basis of the third set of experiments, the CBAM attention mechanism is added, and the mAP is enhanced by 3.2% contrasted with the primitive YOLOv7. The ablation experiment indicates that the enhanced way can preferably enhance the precision of the algorithm without increasing the calculation amount, and all the improvements have a positive effect.

TABLE III
ABLATION EXPERIMENTS ON TT100K

Detect head	CoordConv	CBAM	Epochs	mAP
			300	82.2%
√			300	83.0%
√	√		300	83.8%
√	√		300	84.0%
√		√	300	84.2%
√	√	√	300	85.4%

V. Conclusion

It aims at addressing the problems of low-resolution traffic signs in complex road scenes and missing and false detection of targets. This paper put forward ways and means of a road sign target detection based on improved YOLOv7. In the process of gradually improving the YOLOv7 network structure, the CBAM injection mechanism is added to the connection between the backbone network and the neck network. CBAM is used to enhance the representation ability of neural networks and heighten the feature extraction of traffic sign targets under different road conditions. So as to heighten the awareness about the small target network structure, the small target detection layer has been added in this paper. CoordConv is added before the neck network and detection head in order to the network can more sensitive on spatial position content along the quality route and thus better detect small targets. The abundance of experiments of the improved algorithm on the TT100K datasets have proved the balance between the speed and accuracy of the proposed algorithm. Moreover, ablation experiments can also better illustrate that the proposed YOLOv7-COORD model has better effective detection precision in traffic sign detection and can be superior used in target detection assignments in complex road scenes.

However, the experiment still has some shortcomings in small traffic sign target detection. The complex background and diversity of categories make it impossible for the model to reach the peak of all detection targets at the same time. Future research work and study will continue to adopt advanced algorithms, dataset expansion and annotation, multi-modal information fusion, algorithm optimization, and other strategies. Contionuously enhance the robustness and precision of the YOLOv7 pattern for traffic sign target detection.

REFERENCES

[1] ZHANG M H, "Vehicle Detection Method of Automatic Driving based on Deep Learning," *IAENG International Journal of Computer Science*, vol. 50, no. 1, pp. 86-93, 2023.

[2] ZHANG X Y, GAO H B, ZHAO J H, et al. "Overview of autonomous driving technology based on deep learning." *Journal of Tsinghua University(Natural Science Edition)*, 2018, 58(4): pp. 438-444.

[3] WANG B, CHANG F L, LIU CH SH. "Traffic sign classification based on multi-feature fusion." *Journal of Shandong University (Engineering Science Edition)*, 2016, 46(4): pp. 34-40, 53.

[4] WANG Y, MU CH Y, MA X. "Recognition of traffic signs based on Zernike invariant moments and SVM." *Highway Transportation Science and Technology*, 2015, 32(12): pp. 128-132.

[5] LIANG M J, CUI X Y, SONG Q S,et al. "Traffic sign recognition method based on HOG-Gabor feature fusion and Softmax classifier." *Journal of Traffic and Transportation Engineering*, 2017, 17(3): pp. 151-158.

[6] GU Y L, ZONG X X. "A review of object detection study based on deep learning." *Modern Information Technology*, 2022, 6(11): pp. 76-81. (in Chinese).

[7] GIRSHICK R, DONAHUE J, DARRELL T, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. IEEE, 2014: pp. 580-587.

[8] GIRSHICK. "R. Fast R-CNN." *2015 IEEE International Conference on Computer Vision (ICCV)*. Piscataway: IEEE, 2016: pp. 1440-1448.

[9] REN S Q, HE K M, GIRSHICK R, et al. "Faster RCNN: towards real-time object detection with region proposal networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39 (6): pp. 1137-1149.

[10] HE K M, GKIOXARI G, DOLLÁR P, et al. "Mask RCNN." *2017 IEEE International Conference on Computer Vision (ICCV*). Venice, Italy. IEEE, 2017: pp. 2980-2988.

[11] REDMON J, DIVVALA S, GIRSHICK R, et al. "You only look once: unified, real-time object detection." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2016: pp. 779-788.

[12] LIU W, ANGUELOV D, ERHAN D, et al. "SSD: Single shot multibox detector." *European Conference on Computer Vision*. Cham: Springer, 2016: pp. 21-37.

[13] SUDHA M, GALDIS PUSHPARATHI D V P. "Traffic sign detection and recognition using RGSM and a novel feature extraction method." *Peer-to-Peer Networking and Applications*, 2021, 14(4): pp. 2026-2037.

[14] HAN C, GAO G Y, ZHANG Y. "Real-time small traffic sign detection with revised faster-RCNN." *Multimedia Tools and Applications*, 2019, 78(10): pp. 13263-13278.

[15] WAN J X, DING W, ZHU H L, et al. "An efficient small traffic sign detection method based on YOLOv3." *Journal of Signal Processing Systems*, 2021, 93(8): pp. 899 -911.

[16] YIN J H, QU S J, YAO Z K, et al. "Traffic sign recognition model in haze weather based on YOLOv5." *Journal of Computer Applications*, 2022, 42 (9): pp. 2876-2884.

[17] Wang C Y, Bochkovskiy A, Liao H Y M. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: pp. 7464-7475.

[18] YU J, ZHANG W. "Face mask wearing detection algorithm based on improved YOLO-v4." *Sensors*, 021, 21(9): pp. 263.

[19] SONG Q, LI S, BAI Q, t al. "Object detection method for grasping robot based on improved YOLOv5." *Micromachines*, 2021, 12(11): pp. 1273.

[20] Woo S, Park J, Lee J Y, et al. "CBAM: Convolutional Block Attention Module." *Proceedings of the European Conference on Computer Vision(ECCV)*. 2018: pp. 3-19.

[21] Hu Z F, Wang W H, et al. "Loop Closure Detection Algorithm Based on Attention Mechanism," *IAENG International Journal of Computer Science,* vol.50, no. 2, pp. 592-598, 2023.

[22] WOO S, PARK J, LEE J Y, et al. "CBAM: Convolutional block attention module." *Lecture Notes in Computer Science*, 2018: pp. 3-19.

[23] NIU Z Y, ZHONG G Q, YU H. "A review on the attention mechanism of deep learning." *Neurocomputing*, 2021, 452: pp. 48-62.

[24] Liu R, Lehman J, Molino P, et al. "An intriguing failing of convolutional neural networks and the coordconv solution." *Advances in Neural Information Processing Systems*, 2018, pp. 31.

[25] ZHU Z, LIANG D, ZHANG S H, et al. "Traffic-sign detection and classification in the wild." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2016: pp. 2110-2118.