Journal of Computers Vol. 34 No. 3, June 2023, pp. 63-73
doi: 10.53106/199115992023063403005

Traffic Sign Detection Based on Improved YOLOv5

Hua-Ping Zhou, Chen-Chen Xu*, Ke-Lei Sun

Anhui University of Science and Technology, School of Computer Science and Engineering, Huainan City, Anhui, China
2020201034@aust.edu.cn

Received 13 June 2022; Revised 10 September 2022; Accepted 28 October 2022

Abstract. As a popular research direction in the field of intelligent transportation, various scholars have widely concerned themselves with traffic sign detection However, there are still some key issues that need to be further solved in order to thoroughly apply related technologies to real scenarios, such as the feature extraction scheme of traffic sign images, and the optimal selection of detection methods. For the purpose of overcoming these difficulties. This paper proposes a YOLO-based traffic sign detection framework. Firstly, a lightweight convolution attention mechanism is embedded into the backbone network to obtain the information of space and channel; Secondly, the multi-scale awareness module is used to replace large convolution with 3×3 convolution superposition to improve the receptive field area of the object in the model and enhance the feature fusion performance of the model; Finally, CIoU is used as the loss function of the bounding box to locate the experimental object with high precision. The experimental results show that on the CCTSDB data set, the MAP of this method reaches 91.0%, which is 3.5% higher than the original YOLOv5. Compared with other mainstream object detection algorithms, it has a certain degree of improvement, which proves the effectiveness of this method.

Keywords: traffic sign detection, YOLOv5, object detection, multi-scale

1 Introduction

Traffic signs are one of the key components of a transportation system, because they provide guidance or warning information, such as road conditions and real-time traffic conditions, to vehicles and pedestrians [1]. Full compliance with traffic sign laws can greatly prevent traffic accidents and reduce congestion. It is easy for humans to read traffic signs, however, accurately locating and classifying traffic signs remains a huge challenge for cars. Therefore, traffic sign detection has been concerned by the computer vision community [2].

As a key technology for intelligent transportation, the traffic signs detection algorithm is widely used in self-capable transportation construction and driverless technology. However, the existing algorithm still cannot meet the high standards and real-time requirements of the display requirements. Therefore, improving algorithm efficiency is one of the important research topics in the field of intelligent transportation. At present, the algorithm based on deep learning is the mainstream, and it shows huge potential in the field of traffic signs [22].

Convolutional Neural Network (CNN) has been proven to achieve excellent performance in image classification and object detection. The development of deep learning brings a new direction for object recognition. Therefore, various excellent target detection algorithms have been published one after another. Among them, representative methods can generally be divided into two categories, which are based on proposal methods and unspecified methods [12-15]. Based on the proposed methods include Region-based Convolutional Neural network (R-CNN) series. Based on no-proposed methods include You Only Look Once (YOLO) [3-4] and Single Shot Multibox Detector (SSD) [5]. They run well in the Pascal visual objection detection class and ImageNet large-scale visual identification challenge.

There are many people design some Convolutional Neural Networks which aiming to solve to detect small traffic sign. Dongtao [6] removes the high-rise feature MAP in the original SSD algorithm, adjusts the aspect ration of the low-layer feature, and more anchor frames in the lower layer distribution to enrich the fine features in the traffic sign scenario in the translational characteristics, and MAP is 75.28%. Zhou [11] introduces shallow feature extraction, deep feature extraction, and HyperNet complex feature fusion modules to improve the PVANET network and improve the detection accuracy of the traffic sign, and MAP is 84.1%. In the traffic sign detection task based on video, Cen [7] improved the model of feature generator to improve the accuracy of small object detection, and MAP is 86%. The above methods enhance the feature extraction ability of the model, which

* Corresponding Author

proves that the improved feature extraction ability of the model can effectively improve the detection accuracy and classification effect. And Zhang [21] introduces the DIOU loss function in the return of the border frame to combine with the NMS, ROI and the context information were combined after the ROI Align unified size, thus strengthen object characteristics expression, MAP is 90.20%.

In this paper, an improved YOLOv5 traffic sign detection method is proposed for the detection of traffic sign, which can better detect road traffic signs.

The main contributions of this paper are summarized as follows:

(1) Based on the YOLOv5 architecture and combined with a lightweight Convolutional Block Attention Module (CBAM), the trunk network Module is improved and adjusted to make use of more shallow features.

(2) Design a multi-scale awareness module, enhance model feature extraction capabilities, and obtain more context information.

(3) CIoU is used as the loss function of bounding box regression.

Experiments show that this method has a good effect on small object detection, can detect most traffic signs on the road, and improves the accuracy to a certain extent compared with the original method.

This paper is organized as follows. Section II introduces the related word of YOLO v5 algorithms. Section III the improvement of the original YOLOV5 algorithm this paper. Experimental results and analysis are presented in Section IV. In Section V, the work of the article is summarized, and the limitations of the work and directions for future research presented.

2 YOLOv5 Model

At present, the object detection technology framework based on deep learning is mainly divided into two categories: one-stage and two-stage object detectors. Two-stage detectors first produce a candidate area that may contain objects, and extracts features from each candidate area. Further classification and calibration of the candidate zone is again, resulting in final results, such as Faster R-CNN and Cascade R-CNN ta al. Single stage detector can classify and predict the object directly for each location on the feature map with higher detection efficiency, such as RetinaNet, SSD, CenterNet, etc. Two-stage detectors usually have better detection performance on data sets [11], while single-stage detectors have higher time efficiency on the premise of ensuring detection accuracy, lower requirements on hardware equipment, and better applicability in the field of road traffic sign detection [16].

YOLOv5 is a single-stage object detection algorithm released by Ultralytics. Compared with YOLOv4, YOLOv5 has the characteristics of smaller mean weight file, shorter training time and reasoning speed on the basis of similar detection average accuracy. The network structure of YOLOv5 is divided into four parts: input, backbone, neck and head [17-20].

Input: the input mainly includes Mosaic data enhancement, picture size processing, and three parts: the adaptive anchor frame. Mosaic data enhances four pictures to combine four images to achieve the effect of rich picture background; picture size handles the minimum black edges of different long-width original images, unified zoom as standard size; adaptive anchor box calculation compares the output prediction box with the ground truth box on the basis of the initial anchor box, calculates the gap, then reversely updates, and continuously iterates the parameters to obtain the most appropriate anchor box value.

Backbone: backbone network extracts features of different levels from images through deep convolution operation, mainly using Bottleneck cross-stage local structure Bottleneck CSP and spatial pyramid pooling SPP. The former aims to reduce computation and improve inference speed, while the latter realizes feature extraction of the same feature graph at different scales. It is helpful to improve the detection accuracy.

Neck: neck adopts the structure of combining FPN and PAN, combines the Conventional FPN layer with the bottom-up feature pyramid, integrates the extracted semantic features with location features, and integrates the features of the trunk layer and the detection layer, so as to enable the model to obtain more abundant feature information.

Head: head is the detection structure of YOLOv5.

The YOLOv5 network structure is shown in Fig. 1:

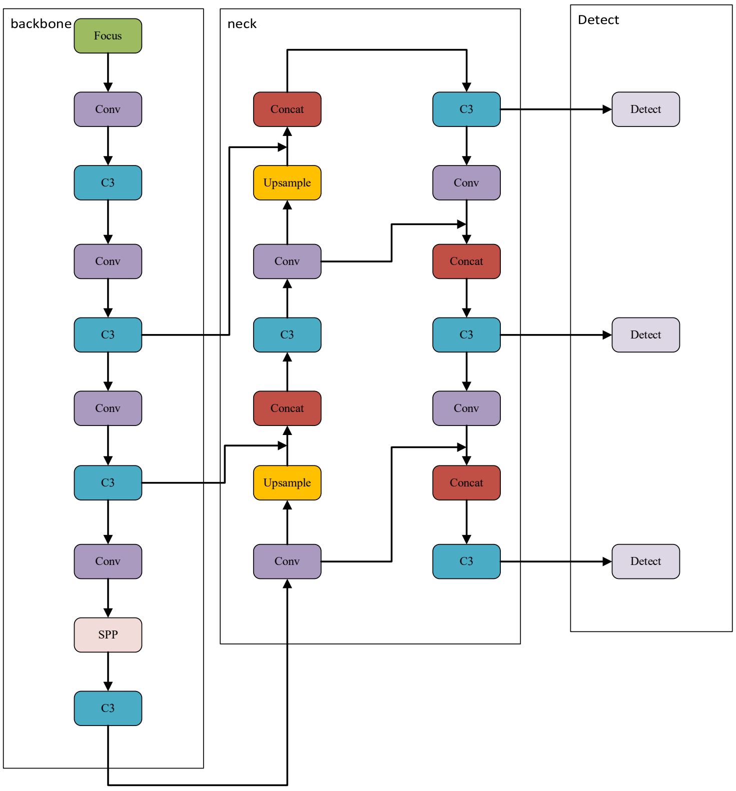

Fig. 1. YOLOv5 network structure

Although YOLOv5 has good detection performance, it still has the following defects for the problems solved in this paper:

(1) Since there are many small goals in the real-world traffic sign, the small object characteristic information is less, the original YOLOV5 algorithm cannot be detected well.

(2) If the receptive field is smaller than or close to the small object scale, the network can't fully play the best effect of small object detection.

(3) When the prediction box is inside the ground truth box and the size of the prediction box is consistent, the original bounding box regression loss function GIoU of YOLOv5 completely degenerated into IoU loss function, which could not achieve high-precision positioning.

3 Improved YOLOv5 Model

For the problem of YOLOv5 in the traffic sign, this paper makes the following improvements to the YOLOv5 original model, increasing the effects of traffic sign detection: (1) A lightweight Convolutional Block Attention Module is introduced into the trunk network to improve spatial and channel perception from the two dimensions of space and channel, and help the model to obtain more context information, so as to locate object information more efficiently. (2) A multi-scale awareness module based on shared convolution superposition is designed to increase the receptive field as much as possible without increasing the number of parameters, so as to obtain information better. (3) CIoU loss function is used to replace the original GIoU loss function to solve the situation that the prediction box is inside the prediction box and the size of the prediction box is consistent, so as to improve the positioning accuracy.

3.1 Introduce CBAM Attention Module

Generally, the common attention module only focuses on the information of channel dimension, which will bring some improvement to the model to a certain extent. However, spatial information is often ignored. The introduction of CBAM attention mechanism can solve this problem well.

CBAM is a simple but effective attention module. It is a lightweight module that can be integrated into most well-known CNN architectures and can be trained in an end-to-end manner. Given a feature map, CBAM deduces an attentional map along two independent dimensions, channel and space, and then multiplies the attentional map with the input feature map to perform adaptive feature refinement. The structure of the CBAM module is shown in the figure. According to the experiment, after integrating CBAM into different models on different classification and detection data sets, the performance of the models has been greatly improved, which proves the effectiveness of this module.

When detecting traffic signs, other useless information is often mixed in. Using CBAM can extract areas of attention, helping the improved YOLOv5 resist confusing information and focus on useful objects.

The structure of CBAM is showed in Fig. 2:

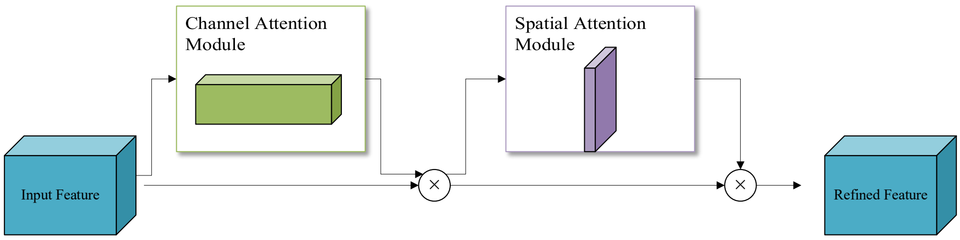

Fig. 2. CBAM structure

CBAM consists of two independent sub modules, channel attention module (CAM) and spatial attention module (SAM), which carry out attention operations in channel and space respectively. This can not only save parameters and computing power, but also ensure that it can be integrated into the YOLOv5 network architecture as a plug and play module.

In this paper, CBAM is inserted after SPP module, and information in channel and space dimensions can be obtained at the same time, which helps the improved model to better capture channel perception and space perception, so as to better locate and identify objects in the region of interest and obtain object information.

3.2 The Multi-Scale Awareness Module

Although the smaller receptive field is conductive to the detection of small object, compared with the object scale, the receptive field less than or close to the small object scale can't give full play to the detection ability of small object [8]. At this point, it is necessary to appropriately increase the receptive field and use the effective context information around the detected object to improve the detection effect of small objects [9].

The Multi-Scale Awareness Module (MSA) expands the receptive field at different scales through multiple shared convolutions and fuses them to introduce more context information. Its design idea comes from Inception Module, in which 1×1, 3×3, 5×5 convolution and 3×3 pooling are stacked together, which increases the width of the network on the one hand. On the other hand, it also increases the adaptability of the original network to multi-scale object characteristics. But unlike Inception Module: The MSA does not directly use 5×5, 7×7 and 9×9 large convolution, but replaces them with two 3×3, three 3×3 and four 3×3 convolution respectively, and the superposition of these multiple 3×3 convolutions can be used as shared convolution. In this way, while introducing context information, it not only alleviates the problem of object resolution reduction caused by multiple convolutions, but also greatly simplifies the network structure, so that MSA can be embedded into our network framework more easily.

The structure of the multi-scale awareness module is shown in Fig. 4, we have four branches for the feature information of the input, which are convolved in different filters, and finally put together in the feature dimension. This structure can utilize a sparse matrix to decompose the principle of intensive matrix calculations to accelerate the convergence speed.

As shown on the left side of Fig. 3: there is a sparse matrix, convolved with a 2×2 matrix, and each element of the sparse matrix needs to be calculated; However, if the sparse matrix can be decomposed into two sub-dense matrices as shown on the right side of Fig. 3, and then convolved with the 2×2 matrix, the local areas with more zeros in the sparse matrix need not be calculated, and the amount of calculation will be greatly reduced. In the MSA module, the decomposition is carried out on the feature dimension. The input data of the traditional convolution layer is convolved with the convolution kernel of one scale, and the data of fixed dimension is outputted. All the output features are basically evenly distributed in this scale range, which is equivalent to output a sparse distributed feature set. However, when the MSA module extracts features at multiple scales (such as 3x3, 5x5, 7×7 and 9×9, etc.), all the output features are no longer evenly distributed, but highly correlated features are clustered together, which can be understood as multiple densely distributed sub-feature sets. In such feature set, irrelevant non-key features are weakened because the features with strong correlation are gathered together. The output of the same number of features is less redundant information. With such a feature set passed layer by layer as input to the reverse calculation, convergence becomes faster [15].

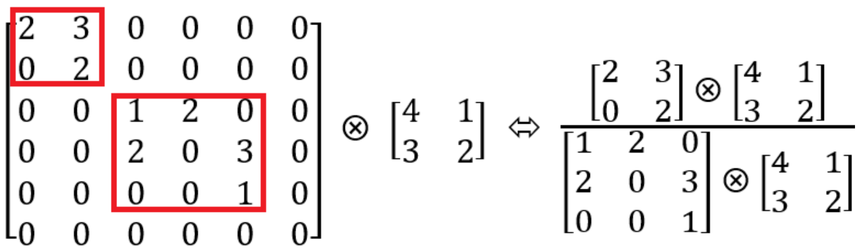

Fig. 3. Sparse matrix is decomposed into sub dense matrix

MSA module first passes 1×1. Adjust the number of channels by convolution, and then analyze the input characteristics respectively. 3×3, 5×5, 7×7, 9×9 convolution operation, and then fusion the receptive fields of different scales. The fusion operation uses “concat”. Finally, classification and bounding box regression are carried out through the output features. The MSA reduces the number of training modules and improves the efficiency of network detection.

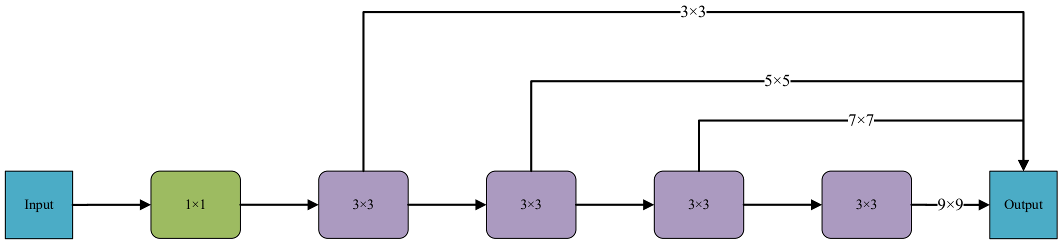

Fig. 4. MSA module structure

3.3 Improved Loss Function

The definition of YOLOv5 loss function is shown in Equation (1):

$$Loss = l_{obj} + l_{cls} + l_{box}, \tag{1}$$

Where l_{obj} stands for confidence loss, l_{cls} stands for classified loss, l_{box} represents the position loss of the ground truth box and bounding box.

In YOLOv5, l_{box} uses the GIoU loss function, as defined in Equation (2):

$$GIoU = IoU - \frac{\left|C-(A\cup B)\right|}{\left|C\right|}, \tag{2}$$

The Equation (2) means: first calculate the minimum closure area of the two boxes, that is, the area of the minimum box containing both the prediction box and the ground truth box, then calculate the IoU, then calculate the proportion of the closure region that does not belong to the two boxes in the closure region, and finally subtract this proportion from IoU to get GIoU.

GIoU loss focuses not only on overlapping regions, but also on non-overlapping regions to solve the problem that the gap between non-overlapping boxes cannot be evaluated. When there is no overlap between the prediction box and the ground truth box, the GIoU decreases as the distance increases and approaches -1. However, there is a problem with GIoU. Once there is an inclusion relationship between the prediction box and the ground truth box, or when the width and height are aligned, the difference set is 0, and GIoU degenerates into IoU, which cannot evaluate the relative position and has slow convergence.

To solve this problem, this paper adopts CIoU as the position loss function of the ground truth box and bounding box, and the calculation equation is shown in Equation (3):

$$CIoU = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \partial v, \tag{3}$$

Among them, the center point of the prediction box is represented by b, the center point of the ground truth box is represented by b^{gt} , ρ represents the Euclidean Distance, c represents the diagonal distance of the smallest rectangle formed by the intersecting prediction box and the ground truth box, ∂ is a balance coefficient, and v represents the parameter of the consistency of length width ratio. The calculation formulas of ∂ and v can be expressed by Equation (4) and Equation (5), where w^{gt} and h^{gt} represent the width and height of the real frame respectively; w and h represent the width and height of the prediction frame, respectively.

$$\partial = \frac{v}{(1-IoU)+v}, \tag{4}$$

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2, \tag{5}$$

The CIoU considers the overlapping area and the center point distance between the prediction box the ground truth box, when the ground truth box parcel prediction box, directly measuring the distance of two boxes, thus considering the bounding box center distance of information and information of the scale of the width to height ratio bounding box, at the same time also takes into account the aspect ratio of the prediction box and the ground truth box aspect ratio, so as to make the border regression result is better, The detection performance of the model is further improved.

This paper makes a series of improvements to the original YOLOv5 algorithm. The improved YOLOv5 structure is shown in Fig. 5:

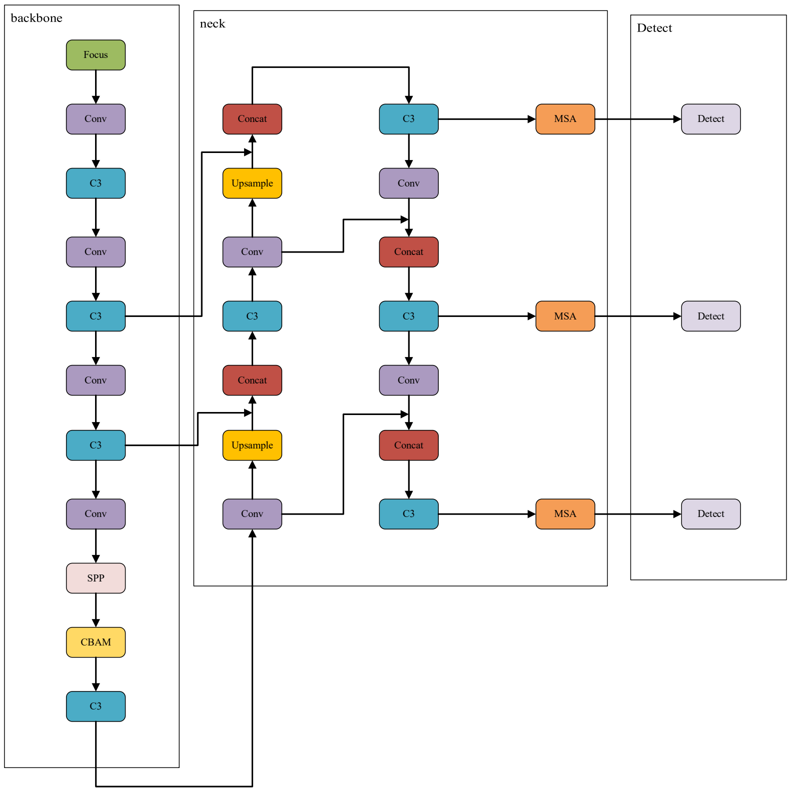

Fig. 5. Improved YOLOv5 structure

4 Experimental Results and Analysis

This paper experiment is based on the Windows10 operating system, NVIDIA GeForce RTX3060 graphics card, and experiments with the model of the Pytorch 1.8.1 deep learning framework.

4.1 The Data Set

This experiment was carried out on CCTSDB Traffic Sign Data set. CSUST Chinese Traffic Sign Detection Benchmark (CCTSDB) [10] is developed by Changsha University of Science and Technology Intelligent Processing Center of Integrated Transportation Big Data, Hunan Provincial Key Laboratory Zhang Jianming's team completed the production. CCTSDB is a traffic sign data benchmark in China. It contains 15,734 images from urban roads and expressways with a resolution of 1000×350 to 1024×768. Data categories are mandatory, Prohibitory, and Warning. In this paper, 4000 images with low similarity in different scenarios were selected from CCTSDB dataset (3000 for training and 1000 for testing).

4.2 The Evaluation Index

In this experiment, Mean Average Precision (MAP) was taken as the average index of the algorithm, and its calculation method was shown in Equation (6):

$$MAP = \int_0^1 P(R)dR, \tag{6}$$

Where P and R represent Recall and Precision respectively, and they can be calculated by Equation (7) and Equation (8) respectively:

$$Recall = \frac{TP}{TP + FN}, \tag{7}$$

$$Precision = \frac{TP}{TP + FP}, \tag{8}$$

Where TP is the number of positive samples divided into positive samples, FN is the number of positive samples incorrectly divided into negative samples, FP is the number of negative samples incorrectly divided into positive samples, $TP+FN$ is the number of all positive samples, and $TP+FP$ is the number of all positive samples divided into positive samples.

4.3 Performance Comparison with Other Methods

In order to verify the effectiveness of the improved algorithm in this paper, we compare it with the current mainstream object detection algorithms, including Faster R-CNN, YOLOv4, SSD and YOLOv5 original algorithms.

Table 1. AP comparison of different algorithms

Algorithms	AP/%		
	mandatory	prohibitory	warning
Faster R-CNN	83.8	80.1	87.8
YOLOv4	86.3	84.1	88.0
SSD	85.0	85.7	90.7
YOLOv5	86.1	86.5	89.9
Ours	89.9	90.7	92.3

Table 1 lists the classification AP values obtained by the current mainstream object detection algorithm and the improved method in this paper on the CCTSDB data set. Among them, the detection accuracy of YOLOv5's three types of traffic signs on the data set are 86.1%, 86.5% and 89.9% respectively, which does not achieve much advantage compared with other commonly used methods. Therefore, this paper improves it. Through the improvement in this paper, the detection accuracy of YOLOv5 in three types of traffic signs are 89.9%, 90.7% and 92.3% respectively, which are 3.8%, 4.2% and 2.4% higher than the original YOLOv5 algorithm. In addition, compared with other algorithms, the accuracy of the improved algorithm is improved to some extent.

Table 2 lists the MAP comparison between the current mainstream target detection algorithm and the improved algorithm in this paper.

Table 2. MAP comparison of different algorithms

Algorithms	MAP/%	FPS
Faster R-CNN	83.9	25.3
YOLOv4	86.1	84.4
SSD	87.1	89.5
YOLOv5	87.5	88.0
MN-YOLO [21]	90.2	85.3
Ours	91.0	87.3

According to the analysis of Table 2, compared with other algorithms, the improved algorithm in this paper also has a certain improvement. In addition, the MAP of the improved algorithm in this paper is 3.5% higher than that of the original YOLOv5 algorithm, and there isn't too much difference between the FPS of the two algorithms, so this shows the effectiveness of the improved method in this paper.

In addition, we draw the P-R curves according to the Precision and Recall values of the above YOLOv5 and the improved YOLOv5. As shown in Fig. 6, the P-R curve detection results of the improved YOLOv5 network in this paper can all surround the other detection model, indicating that the method in this paper can effectively detect traffic signs not detected by YOLOv5.

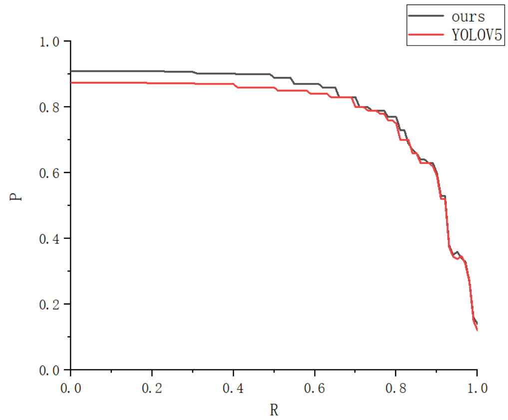

Fig. 6. PR curves of different algorithms

4.4 Ablation Experiment

Compare the original algorithm with the improved algorithm, and the results are shown in Table 3.

Table 3. Comparison of different improved ablation methods

	CBAM	MSA	CIoU	MAP/%
				86.8
YOLOv5	√			87.6
	√	√		89.3
	√	√	√	90.0

According to the analysis of Table 3, the MAP of the original YOLOv5 algorithm is 86.8%. Due to the embedding of CBAM attention module, the model can better obtain the information in spatial and channel dimensions. At this time, the MAP of the algorithm is increased by 0.8%; On this basis, the designed multi-scale sensing module is embedded, and the model obtains more context information. At this time, the MAP of the algorithm is improved by 1.7% compared with the previous one; In addition, on this basis, the position loss function of the ground truth box and bounding box is improved. At this time, the MAP is increased by 0.7%.

4.5 Images Visualization

Some images are randomly selected for visualization. Fig. 7 shows the experimental detection results. From left to right are the detection results of the original YOLOv5 and the improved YOLOv5. It can be seen that the improved method in this paper can detect the category of traffic signs and improve the detection accuracy to a certain extent on the basis of the original algorithm.

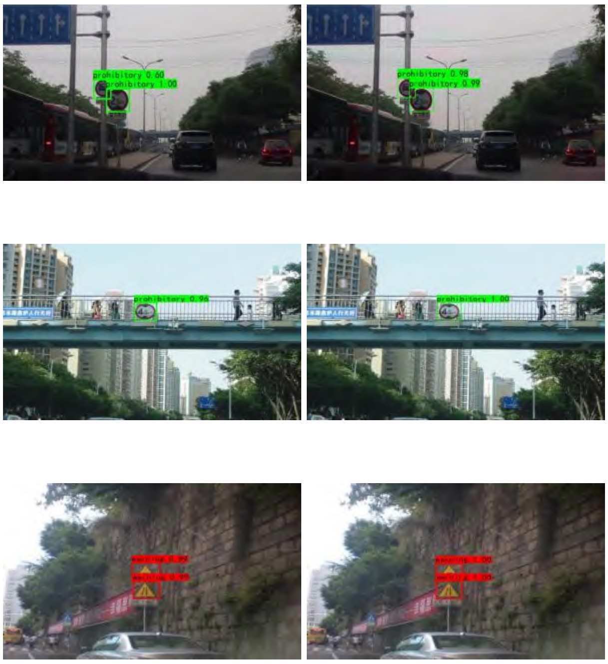

Fig. 7. Some test results of the experiment

5 Conclusion

Based on the YOLO series, a novel traffic sign object detection algorithm is proposed in this paper, to meet the high efficiency requirements of traffic signs detection. Firstly, a lightweight convolution attention mechanism is embedded into the backbone network to obtain the information of space and channel; Secondly, the multi-scale awareness module is used to replace large convolution with 3×3 convolution superposition to improve the receptive field area of the object in the model and enhance the feature fusion performance of the model; Finally, CIoU is used as the loss function of the bounding box to locate the experimental object with high precision. Experimental results prove that the improved algorithm shows high detection accuracy to meet the requirements of detection.

Traffic signs detection is one of the key techniques of intelligent transportation system with high requirements for accuracy and real-time. In the future work, under the premise of maintaining the detection of traffic signs, we will further improve the real-time requirements for the object of traffic signs and help the development of the intelligent transportation system.

References

[1] S. Zhou, X.L. Zhi, D. Liu, H. Ning, L. Jiang, F. Shi, A Convolutional Neural Network-Based Method for Small Traffic Sign Detection, Journal of Tongji University (Natural Science) 47(11)(2019) 1626-1632

[2] H. Liu, Y. Liu, F. Sun, Traffic sign recognition using group sparse coding, Information Sciences 266(2014) 75-89.

[3] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2016.

[4] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, SSD: Single shot MultiBox detector, in: Proc. European conference on computer vision, 2016.

[6] D.T. Zhu, J. Chen, X. Yang, H. Shao, L.L. Su, Traffic sign detection method of improved SSD based on deep learning, in: Proc. IEEE 4th International Conference on Computer and Communications, 2018.

[7] C. Han, G.Y. Gao, Y. Zhang, Real-time small traffic sign detection with revised faster-RCNN, Multimedia Tools and

Applications 78(10)(2019) 13263-13278.
[8] Z.G. Liu, J. Du, F. Tian, J. Wen, MR-CNN: A multi-scale region-based convolutional neural network for small traffic sign recognition, IEEE Access 7(2019) 57120-57128.
[9] Y.K. Chen, Y.W. Li, T. Kong, L. Qi, R.H. Chu, L. Li, J.Y. Jia, Scale-aware Automatic Augmentation for Object Detection, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2021.
[10] J.M. Zhang, Z.P. Xie, J. Sun, X. Zou, J. Wang, A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection, IEEE Access 8(2020) 29742-29754.
[11] G.Z. Xu, Y. Zhou, B. Dong, C.C. Liao, Traffic signs recognition based on improved Cascade R-CNN, Transducer and Microsystem Technologies 40(5)(2021) 142-145,153.
[12] L. Zhao, S. Li, Object detection algorithm based on improved YOLOv3, Electronics 9(3)(2020) 537.
[13] G.Q. Li, L.Z. Xu, Application of local receptive field based broad learning system, Computer Engineering and Applications 56(9)(2020) 162-167.
[14] W.J. Song, M.Y. Fu, Y. Yang, An efficient traffic signs recognition method for autonomous vehicle, Robot 2015(1) (2015) 102-111.
[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2015.
[16] X. Liu, X.K. Jiang, H.C. Hu, R. Ding, H. Li, C.L. Da, Traffic Sign Recognition Algorithm Based on Improved YOLOv5s, in: Proc. 2021 International Conference on Control, Automation and Information Sciences (ICCAI), 2021.
[17] F. Liu, Traffic sign detection based on yolov4 tiny, Information technology and informatization 2021(5)(2021) 18-20.
[18] C. Dewi, R.C. Chen, Y.T. Liu, X.Y. Jiang, K.D. Hartomo, Yolo V4 for Advanced Traffic Sign Recognition With Synthetic Training Data Generated by Various GAN, IEEE Access 9(2021) 97228-97242.
[19] M. Feng, W.Z. Ma, Z.M. Hou, Research on Traffic Sign Detection Algorithm Based on YOLO, in: Proc. AIPR ’21: Proceedings of the 2021 4th International Conference on Artificial Intelligence and Pattern Recognition, 2021.
[20] J.F. Wang, Y. Chen, Z.K. Dong, M.Y. Gao, Improved YOLOv5 network for real-time multi-scale traffic sign detection, Neural Computing and Applications 35(10)(2023) 7853-7865.
[21] D.W. Zhang, X.C. Liu, W. Zhou, Z.H. Chen, Y. Yu, Real-time traffic sign detection algorithm based on improved YOLOv3, Journal of Computer Applications 42(7)(2022) 2219-2226.
[22] T.N. Liu, Z.J. Zhu, Y.Q. Bai, G.L. Liao, Y.X. Chen, YOLO-Based Efficient Vehicle Object Detection, Journal of Computers 33(4)(2022) 69-79.