Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 1, 2023

Improved YOLOv5 Traffic Sign Detection

Liming Zhou *, Zhiren Zhu, Fanrun Meng, Fankai Chen, Chen Liu

College of Electrical Engineering, Tianjin University of Technology and Education, Tianjin, China
*** Corresponding author:** Liming Zhou (Email: 18714557968@163.com)

Abstract: Aiming at the problems such as low accuracy of traffic sign detection and poor real-time performance, a traffic sign detection algorithm based on YOLOv5 is proposed. First, the C2f module is introduced into the backbone network to obtain richer gradient flow information and enhance the feature fusion capability of the target. Second, the SimAM attention module is introduced into the backbone network to enhance the target features and weaken the background features to improve the feature extraction capability of the network model, which in turn improves the detection accuracy of the network model. The experimental results show that compared with the original algorithm, the mAP@0.5 increased by 1.5%, the mAP@0.5:0.95 increased by 2.6%, and the detection speed increased by 28.82%. The improved algorithm detection accuracy can reach 95.5%, and the detection speed can reach 58.82FPS, which can meet the requirements of real-time accurate detection of traffic sign detection.

Keywords: YOLOv5; Traffic Sign Detection; Attention Mechanism.

1. Introduction

Traffic sign detection plays a crucial role in the field of intelligent transportation and automatic driving. In real scenarios, vehicle-mounted cameras often fail to accurately recognize traffic signs due to lighting, weather, distance and other factors, which poses a great potential danger to driving safety. Therefore, it is of great significance to design a traffic sign detection algorithm that takes into account both detection speed and detection accuracy [1].

In recent years, with the continuous development of automatic driving, traffic sign detection has become one of the research hotspots in the field of automatic driving. Traffic sign detection methods are divided into two categories: methods based on color threshold extraction and methods based on deep learning, and methods based on color threshold extraction are susceptible to environmental factors, resulting in poor detection results [2].With the gradual maturity of deep learning algorithms, traffic sign detection methods based on deep learning are beginning to be gradually applied to the field of intelligent transportation and autonomous driving. Tian Zhi et al [3]proposed a traffic sign detection algorithm based on SSD model, which effectively improves the detection effect of the algorithm by introducing low-level feature maps and performing k-means clustering analysis, but its recall rate is only 77.34%.Yang Xiang et al [4]proposed a real-time detection algorithm for small-target traffic signs with improved YOLOv5. The algorithm's detection speed is 46.2 FPS on the TT100K traffic sign dataset, but its detection accuracy is only 87.3%.JW et al [5] proposed an improved YOLOX traffic sign detection method by introducing Swin-Transformer backbone network on the basis of YOLOX, and the improved algorithm improves the detection accuracy by 2.3%, but its detection speed is only 23FPS.Qian Zhang et al [6] a traffic sign detection method based on YOLOv5s, through the introduction of the Transformer convolution module and PANet network structure, while reducing the number of model parameters greatly improves the detection performance of the model, and its detection speed can reach up to 142.86FPS, the detection accuracy is only 78.9%.Han Yanjiang et al [7]proposed an improved traffic sign detection method for YOLOv5, which effectively improves the detection accuracy of traffic signs under different illumination by introducing a coordinate attention mechanism and a bidirectional feature pyramid.

In order to better balance the detection speed and detection accuracy, this thesis proposes a traffic sign detection algorithm based on improved YOLOv5.Aiming at the problem that traffic sign detection is susceptible to the influence of the surrounding environment resulting in inconspicuous target features, the C2f module is first introduced into the backbone network to enhance the feature fusion capability of the target by acquiring richer gradient flow information; Secondly, by introducing the SimAM attention mechanism, the feature extraction ability of traffic signs in images is improved. The experimental results show that the improved model can meet the requirements for traffic sign detection.

2. Base Model

The YOLO series of network models, as a typical representative of single-stage target detection models, can realize end-to-end real-time detection. The YOLOv5[8] network model is further improved on the basis of YOLOv4, which has higher detection accuracy and faster detection speed compared with YOLOv4.The YOLOv5 network model is divided into five different network models according to the width and depth of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5n, and YOLOv5x. In order to ensure the real-time traffic sign detection, this paper chooses YOLOv5s as the baseline model for improvement.

The network structure of YOLOv5s is divided into four parts: input, backbone, neck and head networks. The input adopts Mosaic data enhancement, and the input image is randomly stitched by then scaling, random cropping, and random arrangement, which effectively improves the performance of small target detection while increasing data diversity. The backbone network adopts the Focus structure to slice the image operation, which makes the input channel expand to four times of the original one, and effectively improves the computational power without information loss information. The neck network adopts the structure of FPN+PAN, which effectively enhances the feature fusion capability of the network by combining upsampling and subsampled.

3. Model Optimization

3.1. C2f Module

The C3 module in YOLOv5 uses CSPNet to extract the idea of shunt and combines the residual structure to improve the feature extraction ability of the network model by increasing the network depth and receptive field. The C2f module [9] is further optimized with the help of C3 module and ELAN idea, so that the C2f module has richer gradient flow information while ensuring lightweight, which is used to enhance the feature fusion ability of the network to obtain higher accuracy and more reasonable delay, so some of the C3 modules in YOLOv5 are replaced with C2f modules, and the structure of the C2f module is shown in Figure 1.

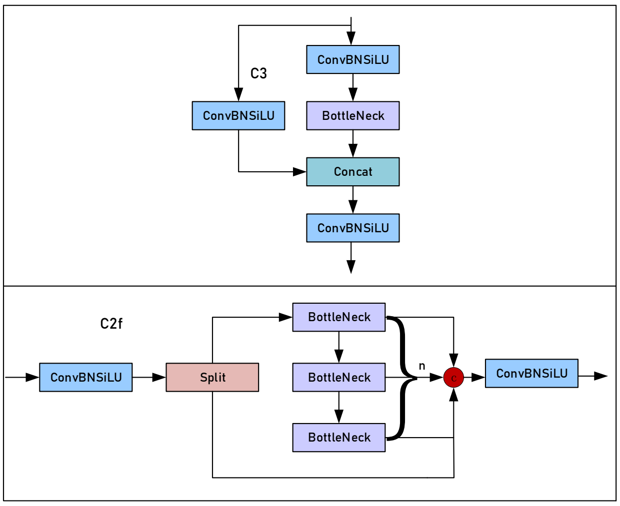


Figure 1. Structural diagram of c3 module and c2f module

3.2. Integration of SimAM Attention Module

Attention mechanism can effectively extract the target features in the image and weaken the background features, the introduction of attention mechanism can effectively improve the performance of target detection. Since traffic sign detection has high requirements for real-time, for this reason the SimAM attention module [10] is introduced to better focus the target features to improve the network performance without introducing additional parameters, and its network structure is shown in Figure 2.

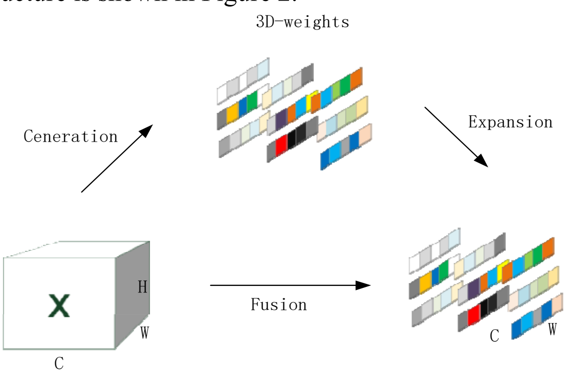


Figure 2. SimAM network structure

SimAM Attention Module is proposed based on neuroscience theory, in neuroscience, the firing pattern of information-rich neurons differs from that of peripheral neurons, and activating neurons usually suppresses peripheral neurons, i.e., space-space suppression. In order to better realize the effective extraction of target features by attention, the neurons with the inhibitory effect in the null domain were chosen to be given a higher importance. The weights of the traffic sign neurons can be obtained by the minimum energy function, which is used to determine the priority between neurons, and the energy formula is shown in equation (1).

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(w_t x_i + b_t))^2 + (1-(w_t t + b_t)) + \lambda w_t^2 \qquad (1)$$

where t and xi are the target neuron and other neurons under the X channel, respectively, M is the number of neurons on that channel, wt is the weight, and bt is the bias. From equation (1), the lower the energy, the more the target neuron distinguishes from other neurons and the higher the importance, according to the definition of the attention mechanism, the feature enhancement formula is shown in equation (2).Replacing all C3 modules in the backbone network with C2f modules and introducing SimAM attention module before the SPPF module, the improved network structure diagram is shown in Figure 3.

$$\tilde{X} = sigmoid(\frac{1}{E}) \odot X \qquad (2)$$

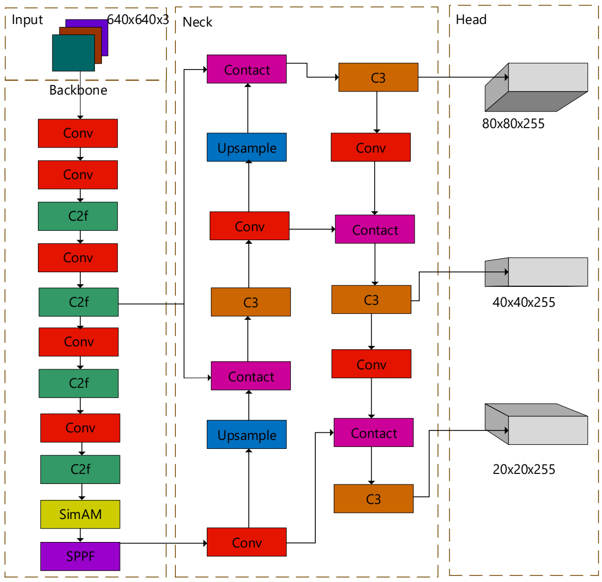


Figure 3. Improved YOLOv5 network structure

4. Experimental Result and Analysis

4.1. Datasets

In this paper, the Chinese traffic sign dataset CCTSDB2021 [11], which has been released by Changsha University of Science and Technology, is used to train, validate and test the network model, the CCTSDB2021 contains 17,856 images, which are categorized according to their meaning as mandatory, prohibition and warning. In this paper, 3270 images are randomly selected from them and the dataset is divided into training set, validation set, and test set according to the ratio of 6:2:2, and its label distribution is shown in Figure 4.

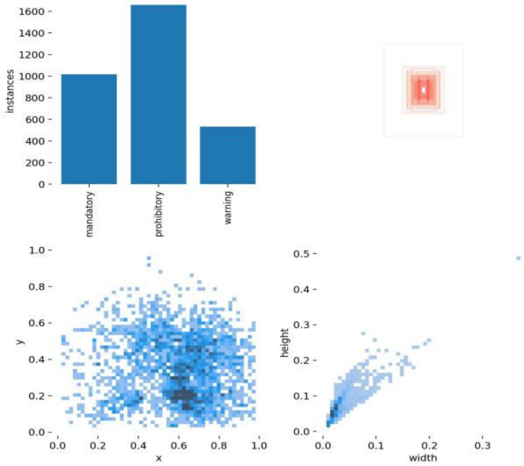


Figure 4. Distribution of CCTSDB2021 labels

4.2. Experimental Environment and Parameter Settings

In this paper, we use YOLOv5s as the experimental benchmark, the input image size is 640x640, batch_size is set to 8, the epochs are 300, all the experiments are carried out in the same environment, the specific experimental environment configuration is shown in Table 1.

Table 1. Experimental environment configuration

Hardware or software	Parameters
GPU	NVIDIA GeForce RTX 3080
CPU	Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
Deep learning frameworks	Pytorch 1.12.1
GPU accelerated environment	CUDA 1.17.1
Programming language	Python3.9

4.3. Evaluation Indicators

This experiment uses Precision (P), Recall (R), mean Average Precision (mAP), Frames Per Second (FPS), where precision indicates the proportion of predicted positive samples to all predicted positive samples, recall indicates the proportion of correctly predicted samples among all samples, mean average precision is used to measure the performance of the model detection, and frames per second is used to detect the number of images that can be detected per second, the specific formula is as follows.

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$mAP = \int_0^1 P(R)d(R) \tag{5}$$

Where TP represents the number of true positive samples, FP represents the number of false positive samples, FN represents the number of false negative samples, and P and R represent accuracy and recall.

4.4. Analysis of the Results

To evaluate the effectiveness of the improved model, the original algorithm is compared with the improved algorithm, and the mAP@0.5 curve comparison plot is shown in Figure 5. as can be seen from Figure 5., the original algorithm and the improved algorithm work well. Compared with the original algorithm, the detection effect of the improved algorithm is significantly improved, which proves that the improved algorithm is effective.

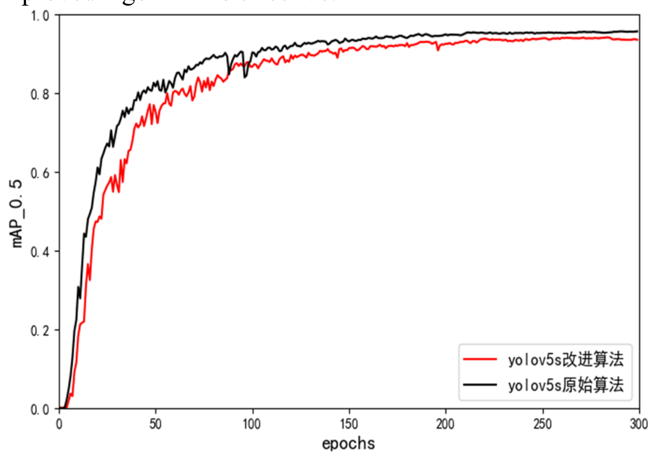


Figure 5. Comparison chart of mAP@0.5 curve

4.5. Ablation Experiments

In order to further verify the reliability of the improved model for improving the accuracy of traffic sign detection, the introduced module and ablation experiment are carried out, and the experimental results are shown in Table 2.As can be seen from Table 2, the introduction of C2f has increased the accuracy rate by 0.5%, the recall rate by 0.4%, the mAP@0.5 by 1.1%, the mAP@0.5:0.95 by 1.3%, and the detection speed to 50FPS.The introduction of SimAM reduced the accuracy rate by 2.3%, the recall rate by 0.7%, the mAP@0.5 by 0.1%, the mAP@0.5:0.95 by 0.2%, and the detection speed to 52.91FPS.By introducing the two modules into the YOLOv5s network model at the same time, the accuracy rate decreased by 0.5%, the recall rate increased by 2.4%, the mAP@0.5 increased by 1.5%, the mAP@0.5:0.95 increased by 2.6%, and the detection speed increased to 58.82FPS, which can meet the requirements of traffic sign detection for detection accuracy and detection speed at the same time.

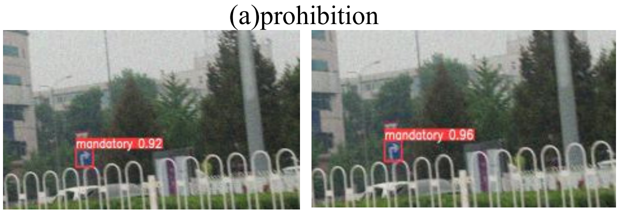
Table 2. Ablation experiments

Algorithm	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	FPS
YOLOv5s	95	88.5	94	66.7	45.66
YOLOv5s-C2f	95.5	88.9	95.1	68	50
YOLOv5s-SimAM	92.7	89.2	94.1	66.9	52.91
YOLOv5s-C2f-SimAM	94.9	90.9	95.5	69.3	58.82

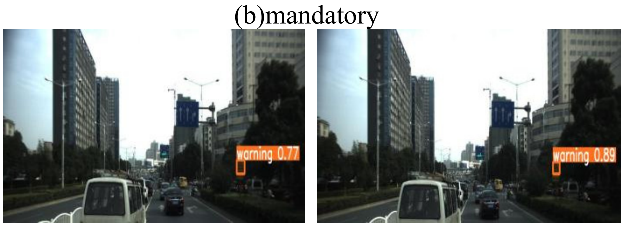
4.6. Detection Effect

In order to prove the effectiveness of the improved model more intuitively, the original model and the improved model are compared and tested, and the pictures are randomly selected from the data set for testing, and the results of the improved model visualization are shown in Figure 6, and the left side is the result of the original model visualization, and the right side is the result of the improved model visualization. The original algorithm in Figure 6(a) mistakenly detects a red street light as a prohibited traffic sign, while the improved algorithm can well detect the difference between a red traffic light and a prohibited traffic sign; Both the original and improved algorithms in Figure 6(b) identify traffic signs in the indicated categories, with the improved algorithm detecting them with higher confidence; In Figure 6(c), both the original algorithm and the improved algorithm can accurately and accurately detect warning traffic signs even if the image light is weak, but the confidence level of warning traffic signs detected by the original algorithm is lower than that of the improved algorithm.

(a)prohibition



(b)mandatory



(c)warning

Figure 6. Comparison of visualization results

5. Conclusion

Aiming at the problem that the target features of traffic signs are not obvious, which leads to low detection accuracy and slow detection speed, this paper proposes an improved traffic sign detection model of YOLOv5, by introducing the C2f module and SimAM attention module in the backbone network to obtain richer gradient flow information and weaken the background features, the feature fusion ability and feature extraction ability of the network model are improved. Experiments show that the improved algorithm can better detect traffic signs. In future research, while ensuring detection accuracy and detection speed, lightweight will be studied to make the model easier to deploy to mobile.

References

[1] LANG Binke,et al. "A traffic sign detection model based on coordinate attention-bidirectional feature pyramid network." Journal of Shenzhen University(Science and Engineering) 40. 03 (2023):335-43.

[2] Zhang Shang, Wang Hengtao,and Ran Xiukang."Lightweight traffic sign detection algorithm based on YOLOv5." Electronic Measurement Technology 45.08(2022):129-135. doi: 10. 19651 / j.cnki.emt.2108726.

[3] Tian Z., Sun Yanyan,and Wei H."Traffic sign detection algorithm based on SSD model." Computer Applications and Software 38.12(2021):201-206.

[4] Yang Xiang, Wang Huabin,and Dong Minggang."Improved YOLOv5's traffic sign detection algorithm." Computer Engineering and Applications 59.13(2023):194-204.

[5] JI Wen,LIU Quanjin,HUANC Chongwen ,et al, YOLOX Trafic Sign Detection Based on Swin-Transformer[J]. Radio Communications Technology ,2023 ,49(3) :547-555.

[6] Zhang, Qian, et al. "Fusion of Transformer and improved PANet for YOLOv5s traffic sign detection."Chinese Journal of Sensors and Actuators36.02(2023):232-241.

[7] Han Yanjiang, Wang Wei,and Wang Fengping."YOLOv5s traffic sign detection method based on coordinate attentionand BifPN."Foreign Electronic Measurement Technology 41.11 (2022): 170-179. doi:10.19652/j.cnki.femt.2204271.

[8] JOCHER G,STOKEN A, BOROVEC J, et al. Utralytics/ YOLOv5: V3.1-bug fixes and performancei-mprovements [EB/0L]. https://doi.org/10.5281/zenodo.4154370, 2020.

[9] CHEN Fankai, and LI Shixin."UAV target detection algorithm with improved yolov5." Computer Engineering and Applications.

[10] Yang, Lingxiao, et al. "Simam: A simple, parameter-free attention module for convolutional neural networks." International conference on machine learning. PMLR, 2021.

[11] Zhang, Jianming, et al. "CCTSDB 2021: a more comprehensive traffic sign detection benchmark." Human-centric Computing and Information Sciences 12 (2022).