applied sciences

MDPI

Article

Traffic Sign Detection Based on the Improved YOLOv5

Rongyun Zhang [1,*], Kunming Zheng [1], Peicheng Shi [1,2], Ye Mei [1], Haoran Li [1] and Tian Qiu [1]

[1] School of Mechanical Engineering, Anhui Polytechnic University, Wuhu 241000, China; 2210120168@stu.ahpu.edu.cn (K.Z.); shipeicheng@126.com (P.S.)
[2] Automotive New Technology Anhui Engineering and Technology Research Center, Anhui Polytechnic University, Wuhu 241000, China
* Correspondence: hanfengzhiwei@163.com

Abstract: With the advancement of intelligent driving technology, researchers are paying more and more attention to the identification of traffic signs. Although a detection method of traffic signs based on color or shape can achieve recognition of large categories of signs such as prohibitions and warnings, the recognition categories are few, and the accuracy is not high. A traffic sign detection algorithm based on color or shape is small in computation and good in real-time, but the color features are greatly affected by light and weather. For the questions raised above, this paper puts forward an improved YOLOv5 method. The method uses the SIoU loss function to take the place of the loss function in the YOLOv5 model, which optimizes the training model, and the convolutional block attention model (CBAM) is fused with the CSP1_3 model in YOLOv5 to form a new CSP1_3CBAM model, which enhances YOLOv5's feature extraction ability and improves the accuracy regarding traffic signs. In addition, the ACONC is introduced as the activation function of YOLOv5, which promotes YOLOv5's generalization ability through adaptive selection of activation by linear–nonlinear switching factors. The research results on the TT100k dataset show that the improved YOLOv5 precision rate increased from 73.2% to 81.9%, an increase of 8.7%; the recall rate increased from 74.2% to 77.2%, an increase of 3.0%; and the mAP increased from 75.7% to 81.9%, an increase of 6.2%. The FPS also increased from 26.88 to 30.42 frames per second. The same training was carried out on the GTSDB traffic sign dataset, and the mAP increased from 90.2% to 92.5%, which indicates that the algorithm has good generalization ability.

Keywords: intelligent driving; traffic sign detection; YOLOv5; attention model; SIoU; ACONC

check for updates

Citation: Zhang, R.; Zheng, K.; Shi, P.; Mei, Y.; Li, H.; Qiu, T. Traffic Sign Detection Based on the Improved YOLOv5. *Appl. Sci.* **2023**, *13*, 9748. https://doi.org/10.3390/app13179748

Academic Editor: Dimitris Mourtzis

Received: 16 July 2023
Revised: 21 August 2023
Accepted: 23 August 2023
Published: 29 August 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the development of intelligent technology, the rapid detection and recognition of traffic signs has become a hot issue for researchers at this stage. Traffic sign detection plays an important role in intelligent driving; after collecting traffic sign information, an intelligent driving system can make the car automatically plan the path and adjust the speed during the driving process, so as to avoid traffic accidents [1–3]. Traffic sign detection is an important part of smart driving which lays the foundation for the analysis and decision-making of environmental perception [4,5]. Accurately detecting traffic signs in complex road environments has become a challenging endeavor. Within the realm of smart driving, traffic sign detection is classified as a vision-based target detection task. This task demands not only high accuracy but also the utilization of limited computational resources, all while maintaining a swift processing rate. As such, it presents a formidable challenge in the development of robust models [6,7]. Traffic sign detection is used to identify traffic signs' locations from video images and detect them, giving them a specific classification [8]. These methods basically use visual messages, for instance, the shape and color of traffic signs, which may lead to poor detection when objects similar to real traffic signs are present in a scene. Although traditional methods can detect traffic signs, the process is tedious, and the ability to extract features is poor, which is not sufficient to

meet the requirements of end-to-end intelligent driving. Therefore, with the continuous advancement of artificial intelligence, convolutional neural networks are widely employed in the field of object recognition.

Currently, some scholars for traffic sign detection tasks have proposed relevant convolutional neural network frameworks [9]. Jin Y. et al. proposed an extraction method for regions of interest which depends on neurons to indicate target information such as path attitude and direction to efficiently obtain traffic sign information from different angles and directions [10]. Hu J. et al. proposed a compact model design, and a squeezing and excitation model was proposed to raise the capability of salient features by learning correlations between two channels [11]. However, direct global processing of information in the channel easily disregards this information. Liang Z. et al. proposed a GIoU loss function based on IoU [12]. Although GIoU achieves a better focus on the detection accuracy of nonoverlapping regions than IoU, the single-frame image detection takes a long time. Tang J. et al. proposed a detection network based on pyramidal multiscale fusion, with a significant improvement in accuracy [13]. However, due to the increase in the number of network parameters, the network is unable to address the performance imbalance between detection accuracy and detection speed. Although this algorithm effectively enhances the precision of object detection, its real-time performance is compromised by the complexity of the network.

In 2015, you only look once (YOLO) was widely utilized in the area of intelligent driving because of its high-speed detection feature, which encapsulates all calculations in the network model and enables YOLO to complete a target detection task at a substantially faster rate. However, its detection accuracy is still unsatisfactory [14]. In the area of traffic sign recognition, Wan J. et al. proposed various optimization strategies based on YOLOv3, which encompassed network pruning, scale prediction branching, and loss function improvement, to address the challenges posed by small traffic sign recognition [15]. In 2020, a one-stage network with both recognition accuracy and speed, YOLOv5, was proposed [16]. YOLOv5 is one-stage target detection algorithm that is mainly made up of an input network, backbone network, neck layer, and output network. The input network uses the image for data enhancement, adaptive anchor frame calculation, and adaptive scaling; the backbone network mainly uses the CSPDarknet53 network, which is employed to extract the rich features information of the image; the neck layer is used for the feature fusion, which is responsible for fusing image information from different scales to obtain more accurate recognition results; the output network is responsible for outputting the detected target information. Ji X. et al. proposed an improved YOLOv5 network model for detecting small targets such as traffic signs; after improving the attention mechanism in the network, small targets can be detected effectively, but there is still some loss in detection accuracy [17].

Therefore, for the problem that detection accuracy and speed cannot be well balanced, this research illustrates a new YOLOv5 algorithm that improves traffic sign recognition. The main contributions of this paper are as follows:

- The CBAM is fused with the CSP1_3 model in YOLOv5 to form a new CSP1_3CBAM model, which not only highlights key features but also suppresses invalid features to achieve effective detection of small target regions.
- The SiLU activation function in the YOLOv5 network is improved, and the ACONC activation function is used, which renders the activation function network more capable of better convergence and improves the robustness of YOLOv5.
- The loss function in the YOLOv5 network is improved, and the SIoU loss function is used, which optimizes the training model to improve its accuracy for small targets.

The remaining portions of this work are broken down into the following sections: The improved YOLOv5 algorithm is introduced in Section 2. The experimental validation and results analysis are introduced in Section 3. The conclusion concludes this article in Section 4.

2. Improved YOLOv5 Algorithm

2.1. Overall Network Framework

Road traffic signs are the guarantor of orderly and safe driving of vehicle drivers on the road, and traffic sign detection in intelligent driving provides corresponding useful information to the whole vehicle control system without interruption. So as to ensure that a intelligent driving system can complete the traffic sign detection task with high precision and high speed, this paper makes some improvements to the YOLOv5 model, which include fusing CBAM into the network, replacing the activation function in the original model with the ACONC activation function, and replacing the loss function in YOLOv5 with the SIOU loss function.

2.2. Attention Mechanism

To tackle the issue of excessive image information in traffic sign detection, as well as the challenge of selecting useful information for the current task objective from a multitude of data, this paper introduces a convolutional attention mechanism into the YOLOv5 network. The addition of the convolutional block attention mechanism can enhance the specific target region of interest, effectively improve the ability of the network to extract features, and weaken irrelevant background regions. The convolutional block attention structure is shown in Figure 1. The convolutional block attention model combines the spatial attention mechanism (SAM) with the channel attention mechanism (CAM). This combination not only reduces the number of network parameters and the computational power needed for network training but also allows for the network's integration into most network architectures [18].

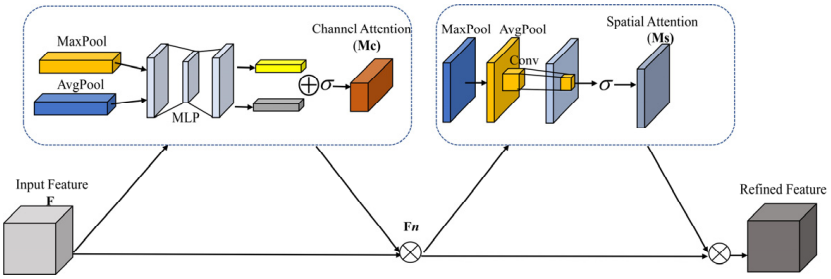


Figure 1. CBAM network structure diagram.

$$M(\mathrm{F}_C) = \sigma(\mathrm{MLP}(\mathrm{F}_{Avg}^{Pool}) + \mathrm{MLP}(\mathrm{F}_{Max}^{Pool})) \tag{1}$$

In Equation (1) above, σ is the sigmoid activation function, MLP is the multilayer perceptron, F_{Max}^{Pool} is the input image F after the MaxPool process, and F_{Avg}^{Pool} is the input image F after the average AvgPool.

The SAM mainly focuses on the position information of the target. The specific process is to obtain two spatial feature maps through the output results of the CAM through AvgPool and MaxPool, splice them according to the channel to obtain an effective feature map, and then perform convolution with a 7×7 convolution kernel and obtain the spatial attention feature map after Sigmoid activation, which is obtained by the spatial attention formula in Equation (2):

$$M(F_S) = \sigma(f^{7\times7}([(F_n{}_{Avg}^{Pool}); (F_n{}_{Max}^{Pool})]) \otimes F_n \tag{2}$$

where σ is the sigmoid activation function, $f^{7\times7}$ is a 7×7 convolution kernel for convolution, and $\otimes$ is element-by-element multiplication.

The CSP1_3 model in YOLOv5 can allow the model to learn more image information, but the model's ability to extract effective information from an input image is weak, and it cannot focus on effective information and suppress invalid information. Therefore, the CBAM is fused with the CSP1_3 model in YOLOv5 to form a new CSP1_3CBAM model.

The CSP1_3CBAM model allows the model to learn more features while inferring the attention map along two different dimensions from the channel and space. Then, the attention map is multiplied with the input feature for self-adaptive feature refinement, so as to highlight the essential information in the input features to achieve the purpose of enhancing the algorithm's ability to extract features and finally raising the detection accuracy for the target. The improved network framework and flowchart are shown in Figures 2 and 3 below.

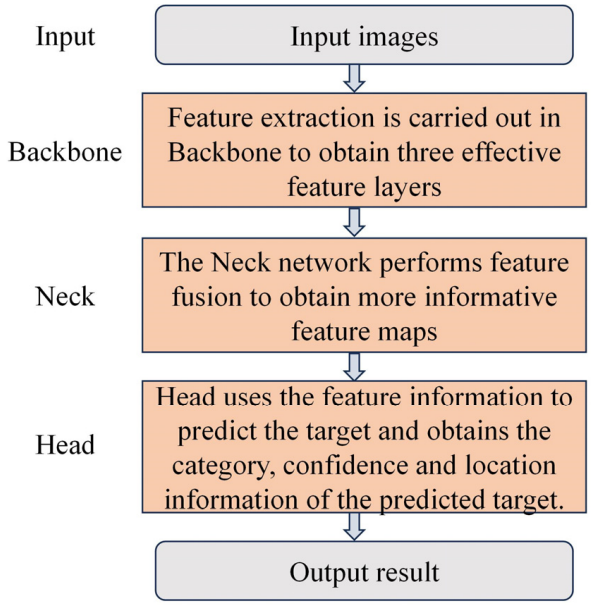


Figure 2. Flowchart of YOLOv5 algorithm.

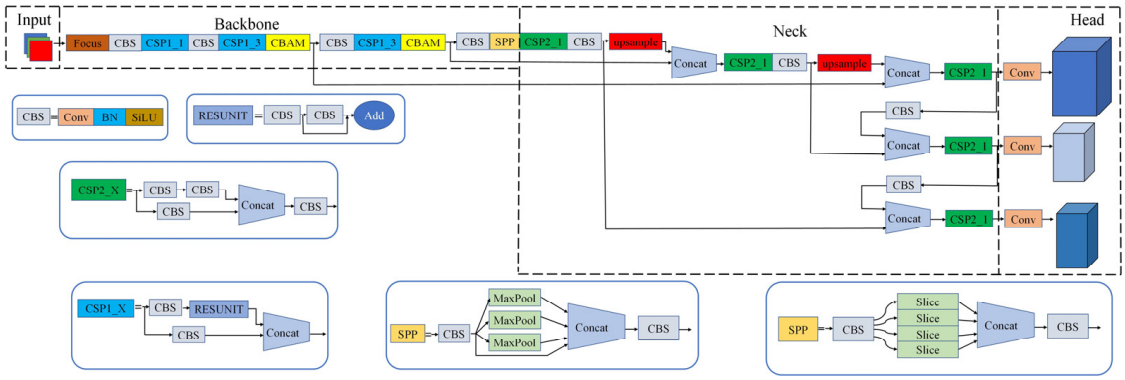


Figure 3. Network framework.

2.3. ACONC Activation Function

In the convolutional neural network structure, the output result of each layer is composed of the result obtained by the input of the upper layer through the linear function, but due to the limited expression space of the linear function, it cannot play a role for nonlinear objects, thus affecting the capability of the neural network. The activation function itself is nonlinear, so the introduction of the activation function into the neural network can make the neural network arbitrarily approach a nonlinear function, enhancing the expressive ability of the model. Thus, choosing a suitable activation function is very important for the model. The ACONC activation function is utilized as the activation function in the model [19]. Compared with the traditional sigmoid activation function, the ACONC activation function decides whether to activate a neuron by introducing a switching factor to learn the parameter switching between nonlinear (activation) and linear

(nonactivation). This activation behavior effectively improves the generalization ability and transmission performance of the model. Its formula is shown in Equation (3).

$$\text{ACONC} = (p_1 - p_2)x \cdot \sigma[\beta(p_1 - p_2)] + p_2 x \tag{3}$$

where s_1 and s_2 denote the parameters that are learned with initial values ($s_1 = 1$ and $s_2 = 0$); σ is the sigmoid activation function; and β is the switching factor.

The switching factor directly determines the degree of the activation function's nonlinearity, and the magnitude of its value depends mainly on the quantity of channels of the model structure (C) and the feature map's size ($H \times W$); its formula is shown in Equation (4).

$$\beta = \sigma \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{c,h,w} \tag{4}$$

2.4. Loss Function Optimization

The network's capacity for prediction will be impacted by the loss function selected. The purpose of the loss function is to calculate the deviation between the predicted value and the real value, calculate the gradient through the loss function, and update the weight in reverse, so a good loss function can greatly optimize the performance of the network. Traditional loss functions depend on the aggregation of the bounding box regression metrics, for instance, the distance, overlapping area, and aspect ratio between the predicted bounding box and the ground truth bounding box. However, most of the currently put-forward and employed methods do not take into consideration the direction matching between the ground truth bounding box and the predicted bounding box. This deficiency gives rise to slow convergence. IoU loss-regresses the four bounding coordinate points of the bounding box as a whole, which is the ratio of the intersection and union between the predicted bounding box and the ground truth bounding box generated by the training process of the network using the merged set [20]. The IoU principle formula and IoU-loss formula are shown in Equations (5) and (6).

$$\text{IoU} = \frac{|M \cap N|}{|M \cup N|} \tag{5}$$

$$\text{IoU-loss} = 1 - \text{IOU} = 1 - \frac{|M \cap N|}{|M \cup N|} \tag{6}$$

where M is the area of the predicted bounding box and N is the area of the ground truth bounding box. In this case, when the IoU-loss value is smaller, it means that the intersection area between two boxes is larger. In contrast, the degree of overlap is low. However, in an actual situation, we may encounter a state in which the predicted bounding box and the true bounding box do not overlap. When the value of IoU is 0 and the value of IoU-loss is 1, the loss function loses its derivative property. The distance between two boxes cannot be calculated, so it cannot continue to learn. Although the size of IoU is the same and can be calculated, the position of the prediction frame is different, which affects the recognition accuracy, and the IoU-loss cannot function properly, as shown in Figure 4 [21].

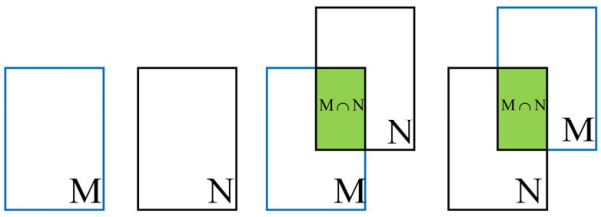


Figure 4. IoU-loss failure scenario.

The relevant loss functions, such as GIoU, DIoU, and CIoU, do not take into consideration the direction mismatch between two boxes, which leads to slow and inefficient convergence of the model. Compared with DIoU-loss, CIoU-loss takes into consideration the aspect ratio in the regression elements but still does not take into consideration the vector angle between two expectations. Therefore, in this article, we choose the SIoU [22] loss function, which takes into consideration the angle of the vector between the two expected regressions and defines a new penalty indicator to improve the speed of training and the accuracy of inference. By introducing the direction into the loss function, the SIoU loss function converges at a faster rate in the training phase with better inference performance compared to existing methods such as CIoU. The proposed improved method improves the model detection accuracy and converges at a faster rate. The SIoU loss function consists of four cost functions: IoU cost, distance cost, angle cost and shape cost. The formula is shown in Formula (7).

$$\text{SIoU-loss} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \tag{7}$$

where IoU is the IoU cost, Δ is the distance loss, and Ω is the shape cost. The formula for the distance cost is shown in (8).

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t}\right) \tag{8}$$

The parameter definition for distance cost is shown in Formulas (9)–(11).

$$\rho_x = \frac{b_{c_x}^{gt} - b_{c_x}}{C_w} \tag{9}$$

$$\rho_y = \frac{b_{c_y}^{gt} - b_{cy}}{C_h} \tag{10}$$

$$\gamma = 2 - \Lambda \tag{11}$$

where C_w and C_h are the width and height of the smallest rectangle bounded by two boxes, γ is the distance value, and Λ is the angle cost.

The formula for the angle cost is as follows:

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t}\right) \tag{12}$$

where $x = \frac{c_h}{\sigma}$, c_h is the center height difference between the true box and predicted box, and σ is the true box and predict box's center distance.

The formula for the shape cost is as follows:

$$\Omega = \sum_{t=w,h} \left(1 - e^{-\omega_t}\right)^{\theta} \tag{13}$$

The parameter definition for the shape cost is shown in Formulas (14) and (15).

$$\omega_w = \frac{|\omega - \omega^{gt}|}{\max(\omega, \omega^{gt})} \tag{14}$$

$$\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \tag{15}$$

where (w, h) and $\left(w^{gt}, h^{gt}\right)$ are the width and height of the two boxes, respectively, θ controls the attention degree of the shape loss, and the value of θ is 4.

3. Experimental Validation and Result Analysis

3.1. Dataset

In this paper, the TT100K dataset and GTSDB dataset are used for training and testing. The TT100K Chinese traffic sign dataset was compiled by Tsinghua and Tencent from video shot by driving recorders. Among the 151 classes in the dataset, only 45 classes have more than 50 instances, which causes a serious imbalance in the data distribution, and training with 151 classes produces overfitting. Therefore, the dataset is filtered, and only 45 classes with more than 50 instances are selected for training. The German Traffic Sign Detection Benchmark (GTSDB), which is widely used to evaluate traffic sign detection, consists of 900 images with a resolution of 1360 × 800 (600 for training and 300 for testing). Figure 5 shows a partial picture of the dataset.



Figure 5. Example of dataset images.

3.2. Evaluation Criteria

In this study, the training is carried out under the Windows 10 operating system using the Pytorch deep learning framework; the programming language Python 3.8; CUDA and CuDNN versions 11.3 and 7.5, respectively; an NVIDIA GeForce GTX3060 graphics card with 6 GB video memory; and a CPU i7-11800H octa-core processor. The size of input images is 640 × 640, the batch size is 16, the optimizer is SGD, and the epoch is 150. Figure 6 shows a comparison of before and after the improvement of the loss function. After 30 rounds, the improved loss function value is significantly smaller than the original model's loss value.

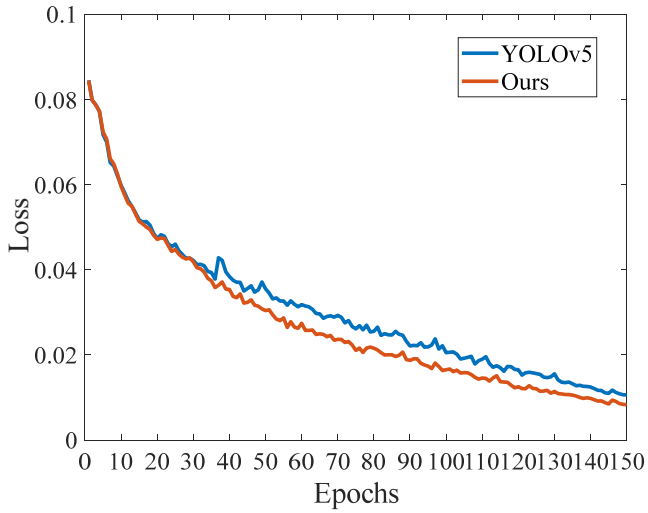


Figure 6. Training effect of loss function after improvement and before improvement.

This paper adopts the precision, the recall, the mean average precision, mAP, and the detection rate, FPS, as the evaluation indices of the performance of the traffic sign target recognition algorithm. TP is the number of traffic signs correctly recognized by the algorithm, FP is the number of positive samples judged by the algorithm that are actually negative samples, and FN is the number of positive samples judged by the algorithm as negative samples. The calculation formulas are expressed as follows:

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{16}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{17}$$

AP is the area under the P–R (precision–recall) curve. mAP is the AP average for all classes. The specific calculation formulas iare as follows:

$$\text{AP} = \int_0^1 \text{P}(r)dr \tag{18}$$

$$\text{mAP} = \frac{\sum_{q-1}^{Q} \text{AP}(q)}{\text{Q}} \tag{19}$$

3.3. Quantitative Analysis

This paper uses the original algorithm and conducts improved algorithm experiments and ablation experiments on the TT100K dataset to test the improvement in model performance from adding the CBAM attention mechanism, improved activation function, and improved loss function. Different improvements are tested in the same hardware environment. Figure 7 shows the mAP before and after the enhancement, indicating a significant improvement in the mAP of YOLOv5.

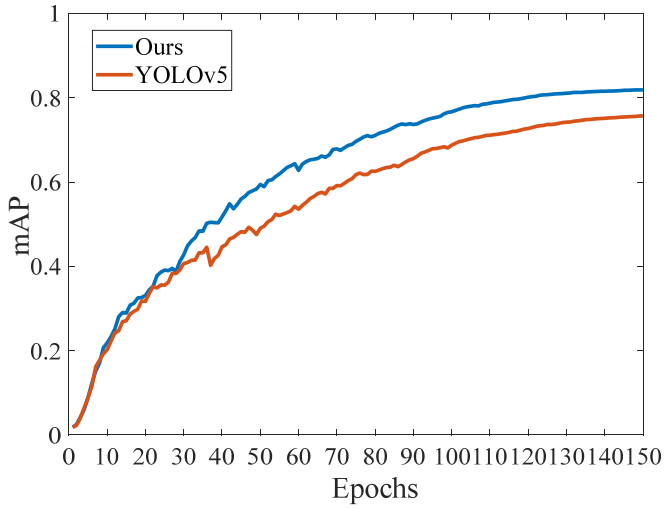


Figure 7. Comparison of model mAP before and after the improvement.

To examine the impact of SIoU, ACONC, and CBAM on the road sign recognition algorithm, a series of ablation experiments were designed and conducted on the TT100K dataset under consistent conditions: (1) YOLOv5 original model; (2) YOLOv5 original model + SIoU; (3) YOLOv5 original model + ACONC; (4) YOLOv5 original model + CBAM; and (5) YOLOv5 original model + SIoU + ACONC + CBAM. The experimental performance of the models is analyzed and compared in Table 1.

Table 1. Ablation experiments.

Model	SIoU	ACONC	CBAM	Parameters	GFLOPs	P/%	R/%	mAP@0.5/%
YOLOv5				7,066,239	16.4	73.2	74.2	75.7
A	√			7,066,239	16.4	79.3	75.8	79.7
B		√		7,469,903	16.7	77.2	74.4	78.7
C			√	7,110,151	16.5	78.0	73.5	78.6
Ours	√	√	√	7,513,815	16.8	81.9	77.2	81.9

When only the loss function is improved, the parameters and GFLOPs do not change, indicating that the number of layers in the network does not change when the loss function is changed; the mAP increases from 75.7% to 79.7%, an increase of 4.0 percentage points. When only the activation function is improved, as the parameters increase, the GFLOPs increase, and the hardware computation efficiency is also higher; the mAP increases from 75.7% to 78.7%, an increase of 3.0%. When only the feature enhancement model is fused, the parameters and GFLOPs increase slightly, the hardware computing efficiency is also increased slightly, and the mAP increases from 75.7% to 78.6%, an increase of 2.9%. When all three are added to the original algorithm, the number of parameters and GFLOPs of the network increase, and the hardware computing efficiency also increases. The improved YOLOv5 precision rate increases from 73.2% to 81.9%, an increase of 8.7%; the recall rate increases from 74.2% to 77.2%, an increase of 3.0%, and the mAP increases from 75.7% to 81.9%, an increase of 6.2%. After calculation, the FPS also increases from 26.88 to 30.42 frames per second. In general, the augmented model's detection speed and accuracy effectively improved.

This paper compares the improved algorithm to other popular algorithms, such as SSD300, Faster R-CNN, TT100K by Zhu [23], YOLOv3, and YOLOv4 [24,25]. The learning rate of all models is set to 0.001; the input size, epoch, and batch size are shown in Table 2 below. The anchor boxes are set to the original data, and the YOLOv5 model adopts the same parameter-setting strategy as us for training and adjusting parameters. The input size is 640 × 640, the batch size is 16, and the epoch is 150. The comparison results are shown in Table 3.

Table 2. Partial training parameters of the model.

Models	Input Size	Learning Rate	Epoch	Batch Size
YOLOv5	640 × 640	0.001	150	16
SSD300	300 × 300	0.001	300	32
Faster R-CNN	416 × 416	0.001	2500	20
Zhu	-	-	-	-
YOLOv3	416 × 416	0.001	200	8
YOLOv4	416 × 416	0.001	200	8
[25]	416 × 416	0.001	200	8
Ours	640 × 640	0.001	150	16

Table 3. Comparison of the improved algorithm with other algorithms.

Models	P/%	R/%	mAP/%	FPS/(f/s)
YOLOv5	73.2	74.2	75.7	26.88
SSD300	76.5	62.3	76.3	14.1
Faster R-CNN	69.8	53.6	80.9	0.6
Zhu	-	-	81.6	5.8
YOLOv3	26.3	37.9	58.1	29.6
YOLOv4	54.0	64.6	74.2	41.7
[25]	-	-	75.2	31.3
Ours	81.9	77.2	82.0	30.42

As shown in Table 3, compared to the SSD300 and Faster R-CNN algorithms, the improved algorithm in this paper is improved more obviously. The mAP of the Faster R-CNN algorithm is 80.9%, the accuracy P is 69.8%, and the recall R is 53.6% when detecting traffic signs in the TT100K dataset. Compared to the Faster R-CNN method, this algorithm has a better detection rate.

3.4. Qualitative Analysis

To illustrate the advantages of the new model presented in this study, four images collected on campus roads, streets, highways, and viaducts were selected for testing. Figure 7 depicts the test findings.

In Figure 8, the leftmost square presents a sample image of traffic signs, while the two squares on the right depict the detection results of the unimproved model and the improved model, respectively. Each prediction result includes three pieces of information: the location, category, and confidence level of the traffic sign. When a scene with multiple targets to be detected is encountered, YOLOv5 is not accurate in detecting some targets. For example, as shown in Figure 8, the target loss detection problem or incorrect detection problem occurs when there are many targets in the input image. In Figure 8a, the no-passing sign is missed because it is neither trained on nor recognized due to an insufficient quantity of instances in the dataset. In Figure 8b, a speed limit sign recognition error occurs, and the front wheel of the bicycle is mistakenly detected as a speed limit sign. In Figure 8c, the confidence level of the initial model detection is 0.85, and the confidence level of the improved model detection is 0.95. In Figure 8d, the improved algorithm in this paper can correctly detect traffic signs with long distances and small targets, and the confidence level is higher than that of the initial model. The new YOLOv5 model employed in this study accurately identifies the traffic sign categories present in the collected images of the dataset without any instances of missed or incorrect detections.

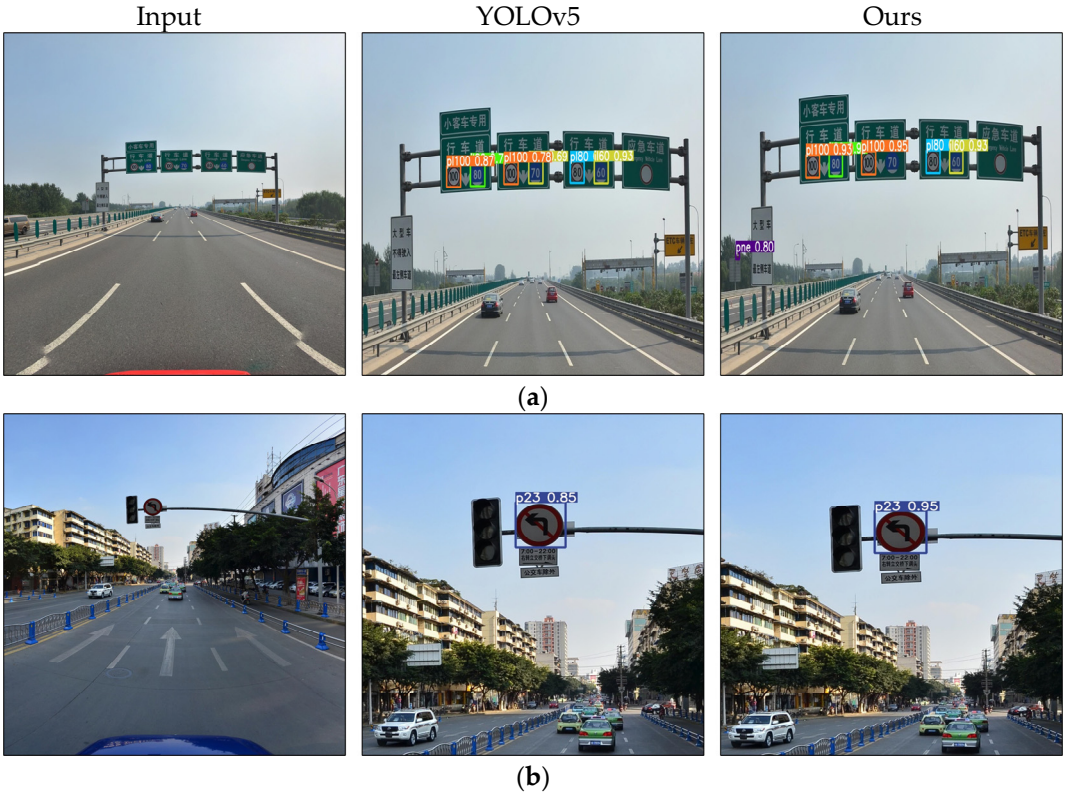
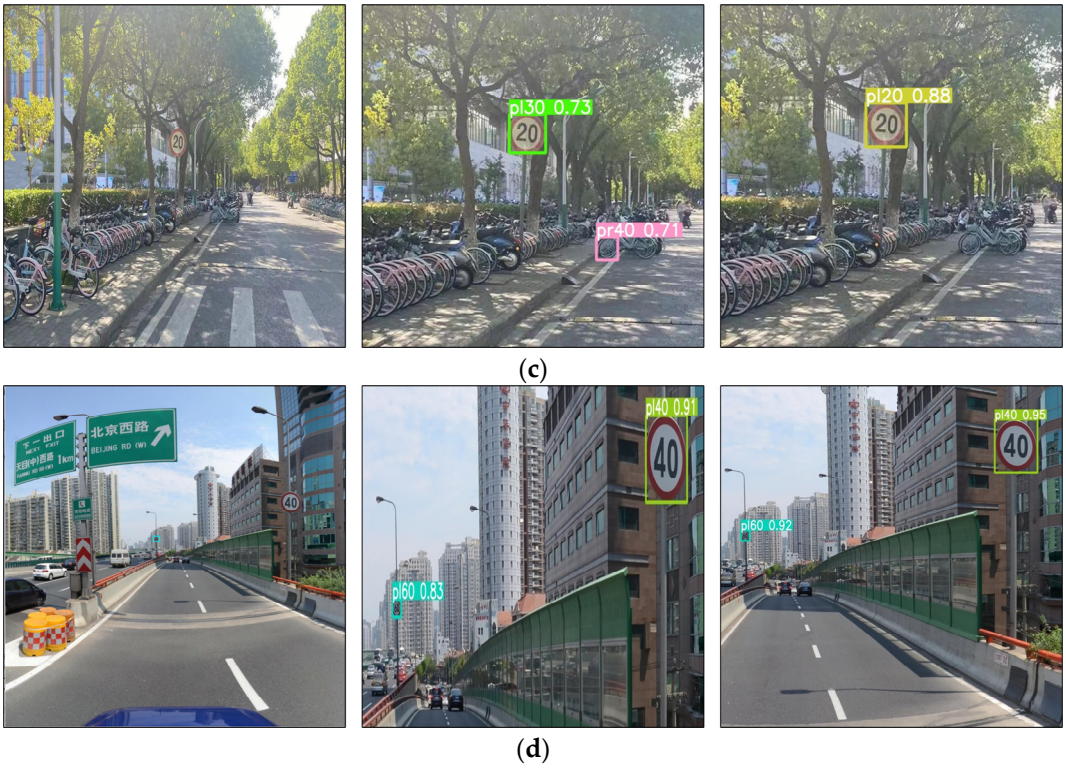


Figure 8. *Cont.*



(**c**)

(**d**)

Figure 8. Improved YOLOv5 algorithm comparison detection graph: (**a**) multiple traffic signs and small targets on a highway; (**b**) single traffic sign on a street; (**c**) single traffic sign on a campus; and (**d**) small and middle targets.

4. Conclusions

Aiming at the problem that current traffic sign recognition technology cannot meet the real-time and accuracy standards required by intelligent assisted driving systems, an improved YOLOv5 algorithm is proposed. Training and testing are performed on the TT100K and GSTDB datasets. The experimental results on the TT100k dataset show that the loss function value of the YOLOv5 algorithm after the improvement is significantly reduced compared to before. The convolutional block attention model and the CSP1_3 model are fused together to create a novel CSP1_3CBAM model. ACONC is selected as the activation function of YOLOv5. Thus, the feature extraction ability of the model for useful information is improved, the recall rate, accuracy rate, detection accuracy rate, and detection speed are improved, and false detections and missing detections during real-time detection are significantly reduced. At the same time, the resulting mAP after training on the GTSDB dataset is 91.5%, indicating that the model has good generalization ability. The light weight of the model will be studied in the future for easy deployment to mobile devices.

Author Contributions: Methodology, R.Z.; software, K.Z.; validation, K.Z., R.Z. and P.S.; investigation, Y.M.; resources, R.Z.; writing—original draft preparation, R.Z.; writing—review and editing, K.Z.; visualization, R.Z.; supervision, H.L.; project administration, T.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Natural Science Foundation of China, Nos. 51605003 and 51575001, the Natural Science Research Project of Anhui Province, China, No. KJ2020A0358, and the young and middle-aged Top Talent Training Program of Anhui Polytechnic University.

Data Availability Statement: No new data were created in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, H.; Wang, K.; Cai, Y.; Liu, Z.; Chen, L. Traffic Sign Recognition Based on improved Cascaded Convolutional Neural Network. *Automot. Eng.* **2020**, *42*, 1256–1262+1269.
2. Liu, Y.; Shi, G.; Li, Y.; Zhao, Z. M-YOLO: Traffic Sign Detection Algorithm Applicable to Complex Scenarios. *Symmetry* **2022**, *14*, 952. [CrossRef]
3. Wang, B.; Han, Y.; Wang, S.; Tian, D.; Cai, M.; Liu, M.; Wang, L. A Review of Intelligent Connected Vehicle Cooperative Driving Development. *Mathematics* **2022**, *10*, 3635. [CrossRef]
4. Gao, T.; Xing, K.; Liu, Z.; Chen, T.; Yang, Z.; Li, Y. Traffic Sign Detection Algorithm based on Pyramid Multi-scale Fusion. *J. Traffic Transp. Eng.* **2022**, *22*, 210–224.
5. Ouyang, Z.; Niu, J.; Ren, T.; Li, Y.; Cui, J.; Wu, J. MBBNet: An edge IoT computing-based traffic light detection solution for autonomous bus. *J. Syst. Archit.* **2020**, *109*, 101835. [CrossRef]
6. Zhu, Y.; Zhang, C.; Zhou, D.; Wang, X.; Bai, X.; Liu, W. Traffic sign detection and recognition using fully convolutional network guided proposals. *Neuro Comput.* **2016**, *214*, 758–766. [CrossRef]
7. Zhu, Y.; Yan, W. Traffic sign recognition based on deep learning. *Multimed. Tools Appl.* **2022**, *81*, 17779–17791. [CrossRef]
8. Huang, Z.; Yu, Y.; Gu, J.; Liu, H. An efficient method for traffic sign recognition based on extreme learning machine. *IEEE Trans. Cybern.* **2016**, *47*, 920–933. [CrossRef] [PubMed]
9. Tian, Y.; Gelernter, J.; Wang, X.; Li, J.; Yu, Y. Traffic sign detection using a multi-scale recurrent attention network. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4466–4475. [CrossRef]
10. Jin, Y.; Fu, Y.; Wang, W.; Guo, J.; Ren, C.; Xiang, X. Multi-feature fusion and enhancement single shot detector for traffic sign recognition. *IEEE Access* **2020**, *8*, 38931–38940. [CrossRef]
11. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef] [PubMed]
12. Liang, Z.; Shao, J.; Zhang, D. Traffic sign detection and recognition based on pyramidal convolutional networks. *Neural Comput. Appl.* **2020**, *32*, 6533–6543. [CrossRef]
13. Tang, J. Detect Lane Line Based on Bi-directional Feature Pyramid Network. In Proceedings of the 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), Guangzhou, China, 5–7 August 2022.
14. Joseph, R.; Ali, F. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
15. Wan, J.; Ding, W.; Zhu, H.; Xia, M.; Huang, Z.; Tian, L.; Zhu, Y.; Wang, H. An Efficient Small Traffic Sign Detection Method Based on YOLOv3. *J. Sign. Process Syst.* **2021**, *93*, 899–911. [CrossRef]
16. Ultralytics.YOLOv5[EB/OL]. Available online: https://github.com/ultralytics/yolov5 (accessed on 10 April 2022).
17. Ji, X.; Lai, C.; Zhou, G.; Dong, Z.; Qi, D.; Lai, L. A Flexible Memristor Model With Electronic Resistive Switching Memory Behavior and Its Application in Spiking Neural Network. *IEEE Trans. Nano Biosci.* **2023**, *22*, 52–62. [CrossRef]
18. Woo, S.; Park, J.; Lee, J.; Kweon, I. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, 8–14 September 2018.
19. Ma, N.; Zhang, X.; Liu, M.; Sun, J. Activate or Not: Learning Customized Activation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
20. Xu, H.; Li, B.; Zhong, F. Light-YOLOv5: A Lightweight Algorithm for Improved YOLOv5 in Complex Fire Scenarios. *Appl. Sci.* **2022**, *12*, 12312. [CrossRef]
21. Wu, S.; Yang, J.; Wang, X.; Li, X. IoU-Balanced loss functions for single-stage object detection. *Pattern Recognit. Lett.* **2022**, *156*, 96–103. [CrossRef]
22. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
23. Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-Sign Detection and Classification in the Wild. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
24. Bochkovskiy, A.; Wang, C.; Liao, H.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
25. Ma, L.; Wu, Q.; Zhan, Y.; Liu, B.; Wang, X. Traffic Sign Detection Based on Improved YOLOv3 in Foggy Environment. In Proceedings of the 2021 International Conference on Wireless Communications, Networking and Applications, Hangzhou, China, 13–15 August 2022; Springer: Singapore, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.