International Core Journal of Engineering Volume 10 Issue 5, 2024
ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

Improved Traffic Sign Recognition Algorithm for YOLOv5

Yunhang Wang*
Tongji University, Shanghai 201804, China
*yunhang_wang@tongji.edu.cn

Abstract

In the fields of autonomous driving and assisted driving, the detection and recognition of traffic signs are crucial for ensuring driving safety. However, challenges such as small target size and diverse dimensions of traffic signs in real-world scenarios make it difficult to obtain relevant feature information, leading to susceptibility to interference from complex backgrounds. In this study, an improved method for traffic sign detection and recognition is proposed based on the YOLOv5s network. To address issues in current traffic sign recognition algorithms, particularly in the identification of small targets in complex backgrounds, a coordinate attention mechanism is embedded into the C3 layer of the original model's neck network. This enhances the focusing capability on crucial areas and effectively eliminates background noise interference. Additionally, to overcome limitations in feature fusion when dealing with variably sized traffic signs in image processing, a Weighted Feature Fusion Network algorithm is introduced. This algorithm weightsly fuses shallow feature maps of the same size containing rich semantic information from the main network with the deep layers of the intermediate detection layer. The ordinary connections in the neck network's middle position are replaced with weighted connections to enhance the multi-size feature fusion capability. Experimental results demonstrate that the improved algorithm, compared to the original YOLOv5s method, achieves a 3.2% improvement in both precision and recall on the CCTSDB 2021 traffic sign detection dataset. The mean average precision is increased by 3.83%, while the detection speed remains at 98.04 frames per second. Consequently, the proposed enhanced algorithm effectively improves the accuracy of traffic sign detection and recognition while maintaining a considerable detection speed.

Keywords

Traffic Sign Recognition; YOLOv5s; Coordinate Attention; Weighted Feature Fusion.

1. Introduction

The task of traffic sign detection aims to accurately identify the categories and positions of traffic signs in images or videos. This technology finds wide applications in fields such as autonomous driving and driver assistance, constituting a vital component of intelligent transportation systems. The precise detection of traffic signs holds significant research significance and practical value for enhancing traffic safety and efficiency. However, owing to the influence of collection conditions in real-world environments, traffic signs often exhibit characteristics such as small dimensions, complex backgrounds, and diverse size variations, leading to recognition difficulties and potential misjudgments that could severely impact vehicle safety. Therefore, real-time and rapid identification of traffic signs remain focal points and challenges in traffic-related tasks.

Both traditional methods and deep learning-based methods have been extensively studied for traffic sign detection and recognition, achieving certain successes. Traditional detection methods primarily

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

utilize color, shape features, or their fusion for traffic sign detection [1-2]. However, traditional traffic sign detection methods heavily rely on manual feature extraction, exhibiting poor robustness and failing to meet the requirements for real-time and accurate traffic sign detection.

In 2014, R-CNN first applied convolutional neural networks to object detection [4]. Since then, deep learning-based methods have been widely adopted for traffic sign detection. Deep learning-based object detection algorithms mainly fall into two categories: two-stage detection and one-stage detection. Two-stage detection algorithms include R-CNN, Fast R-CNN [5], Faster R-CNN [6], etc. One-stage detection algorithms include the YOLO series [7-10] and SSD [11-12], depicting the development status of object detection, as shown in Figure 1.

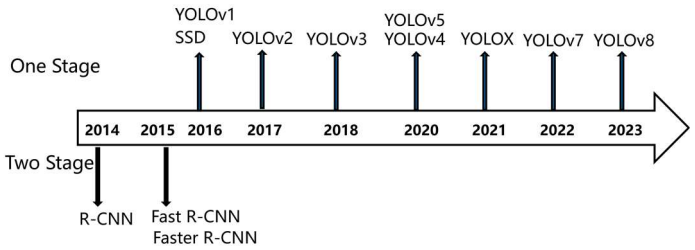


Figure 1. History of object detection development

Yang et al. [13] introduced an attention network based on the Faster R-CNN framework to quickly search all possible regions of interest and roughly classify them into three categories based on color features. Then, another region extraction network generates the final candidate regions from a set of anchor boxes. Jin et al. [14] introduced a feature fusion layer on the basis of the SSD framework and added SE modules to the feature extraction layer after fusion, greatly improving the recognition rate of small traffic signs. The YOLO algorithm was first proposed in 2016, with its key innovation being the use of a grid of S×S cells instead of traditional regions of interest, allowing for the simultaneous output of target coordinates and class probabilities through a single regression. Zhang et al. [15] proposed an improved YOLOv2 algorithm for traffic sign recognition. To reduce computational complexity, they introduced multiple 1×1 convolutional layers in the middle layers of the network and reduced the top convolutional layers. Additionally, to detect small traffic signs, they adopted a denser grid division of the image to obtain finer feature maps. Sichkar et al. [26] first used YOLOv3 to divide traffic signs into 4 categories based on shape for localization and then used another convolutional neural network for accurate classification of the localized traffic signs. On the GTSRB dataset, the experimental results achieved an average precision mAP of 97.22%. Liu et al. [17] introduced MobileNetV2 as the backbone network in the YOLOv5 network, reducing the parameter count of the entire model by 60% while increasing mAP by 0.13%.

Due to the high proportion of small targets and the complexity and diversity of backgrounds and sizes in traffic sign detection, this study aims to improve the accuracy and processing speed of traffic sign detection, thereby enhancing the overall detection performance and making it more suitable for practical engineering applications. The algorithm proposed in this article is based on the YOLOv5s framework and is improved following criteria outlined below:

To address the problem of poor small target recognition by current traffic sign recognition algorithms in complex backgrounds, a traffic sign detection and recognition algorithm based on attention mechanism is proposed. The Coordinate Attention (CA) mechanism is embedded in the C3 structure of the original neck network. By embedding the position information in the channel attention, the focal ability of the critical areas is enhanced, effectively eliminating background noise interference.

To address the limited feature fusion of current object detection algorithms in processing traffic signs of varying sizes in images, a Weighted Feature Fusion (WPANet) network algorithm is proposed. This algorithm combines shallow feature maps of the same size that contain rich semantic information

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

in the main network with medium detection layers in the deep network in an weighted manner. Moreover, the ordinary connections at the middle position of the neck network are replaced with weighted connections to strengthen the ability to fuse multi-size features, making it suitable for traffic signs of different sizes.

2. Overview of YOLOv5s

YOLOv5 is a one-stage detection algorithm that exhibits excellent real-time performance compared to two-stage algorithms. In addition, YOLOv5 demonstrates high detection accuracy, especially in small object detection, among one-stage detection algorithms. This algorithm provides four network models: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. YOLOv5s has the simplest network structure and the fastest object detection speed, while the subsequent three models gradually increase the depth and width of the network, improving accuracy but reducing detection speed. Therefore, in this study, YOLOv5s algorithm is chosen as the base model for traffic sign detection.

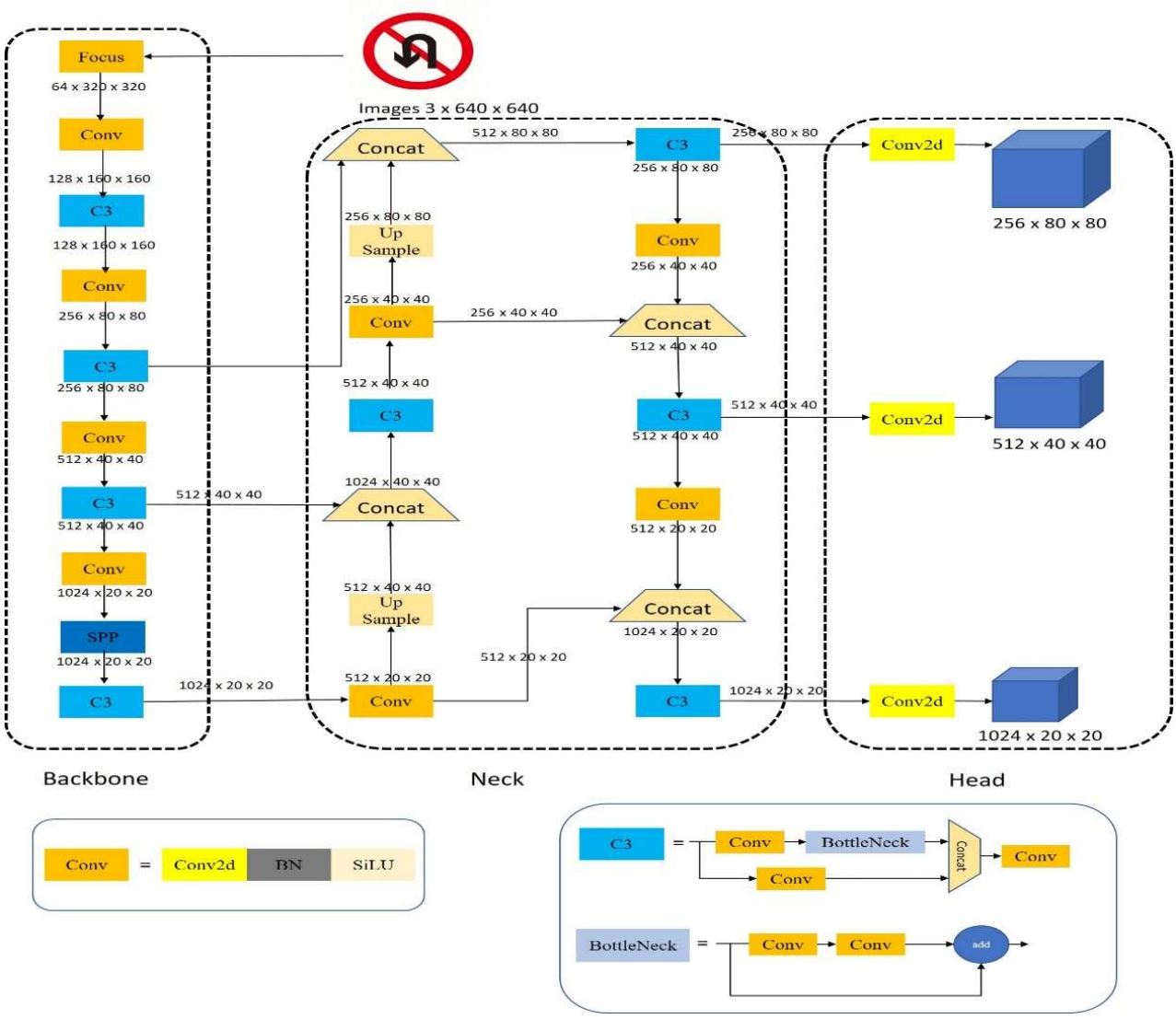


Figure 2. Structure of YOLOv5s network

The YOLOv5s network model mainly consists of four aspects: input preprocessing, backbone network, neck network, and output processing. The input preprocessing part adopts adaptive image padding and Mosaic data augmentation. Mosaic data augmentation randomly crops, rotates, and scales four images and combines them into one image. This data augmentation method effectively

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

improves the accuracy of small object detection while reducing training resource consumption. Furthermore, adaptive anchor box design is used on the input side to better address the varying sizes of targets in different datasets and provide more suitable predefined anchor boxes for the dataset. The backbone network is mainly composed of the Focus structure [19] and the C3 module. The slicing operation of the Focus structure reduces the image resolution without losing image information, while the C3 module consists of a series of residual modules that efficiently reduce computation. The neck network utilizes a Path Aggregation Network [21] to achieve accurate localization and recognition of objects at multiple scales. The three branches of the output processing are used for the detection output of small, medium, and large objects. The network of YOLOv5s is illustrated in Figure 2.

3. Improved based on YOLOv5s

The improved YOLOv5 algorithm has been updated based on the original algorithm. It embeds the coordinate attention mechanism in the C3 structure of the neck network and constructs a feature enhancement module, C3CA. It merges the shallow feature maps with rich semantic information at the same size in the backbone network with the medium detection layers at deeper levels using weighted fusion. It replaces the ordinary connection in the middle position of the neck network with a weighted connection to enhance the feature fusion ability across multiple scales. The structure of the improved YOLOv5s algorithm is shown in Figure 3.

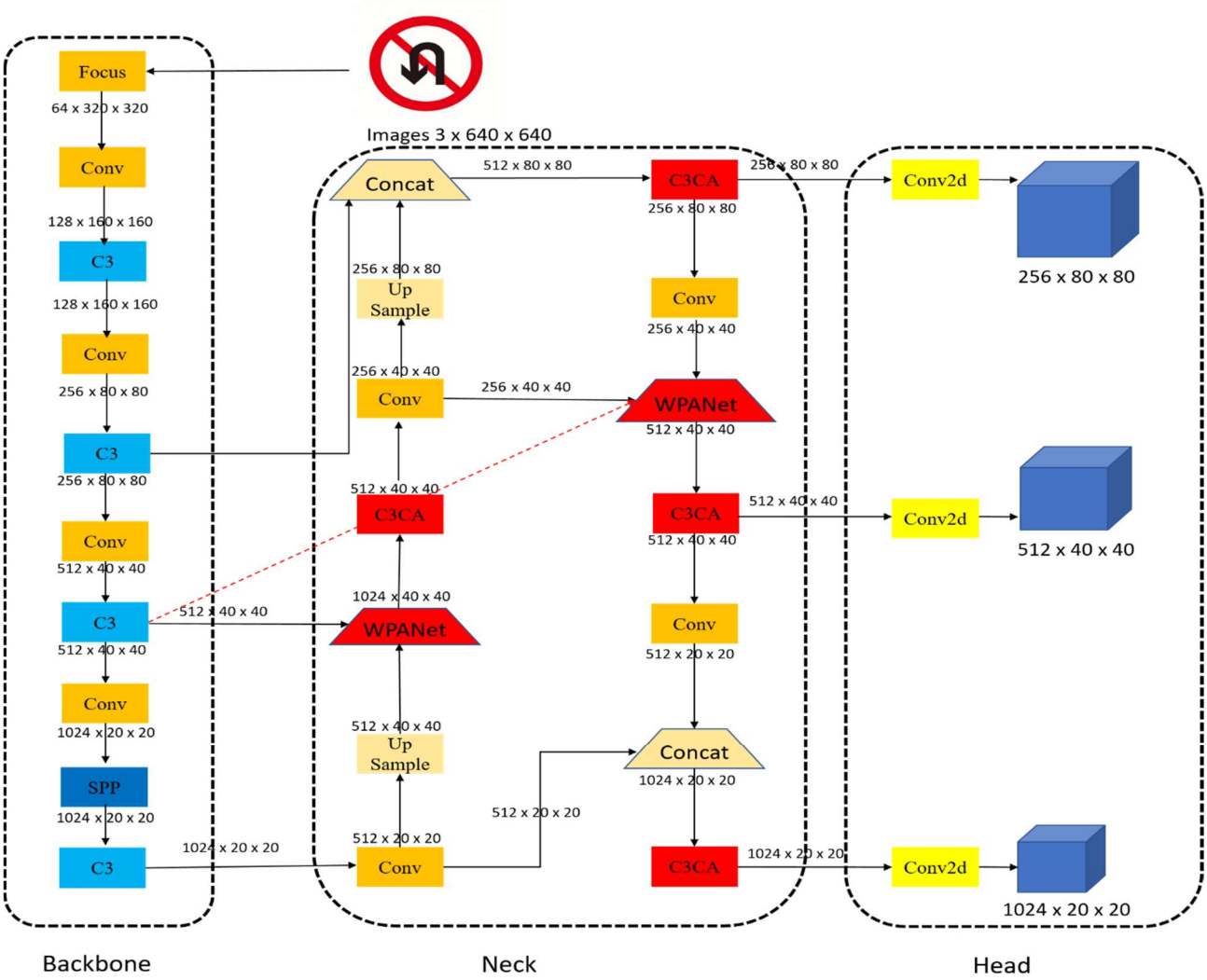


Figure 3. Structure of improved YOLOv5s network

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

3.1 Embedded Attention Mechanism

The attention mechanism is one of the most used concepts in the field of deep learning. Its design inspiration comes from the human visual nervous system, which focuses on important target information during the information processing and suppresses irrelevant background influences. In neural network learning, attention mechanisms are an important component of improving learning performance. The key link in feature extraction of traffic signs lies in the neck network. Therefore, this paper embeds the coordinate attention module CA into the neck part of the network.

The structure of the CA attention mechanism is shown in Figure 4. For any input feature X, first, a pooling kernel with a size of (H, 1) or (1, W) is used to encode each channel along the horizontal and vertical directions. Through the transformations of equations 1 and 2, the generated feature layers possess coordinate information and maintain long-range correlations.

$$Z_c^h(h) = \frac{1}{W}\sum_{0 \leqslant i \leqslant W} x_c(h, i) \tag{1}$$

$$Z_c^w(w) = \frac{1}{H}\sum_{0 \leqslant j \leqslant H} x_c(j, w) \tag{2}$$

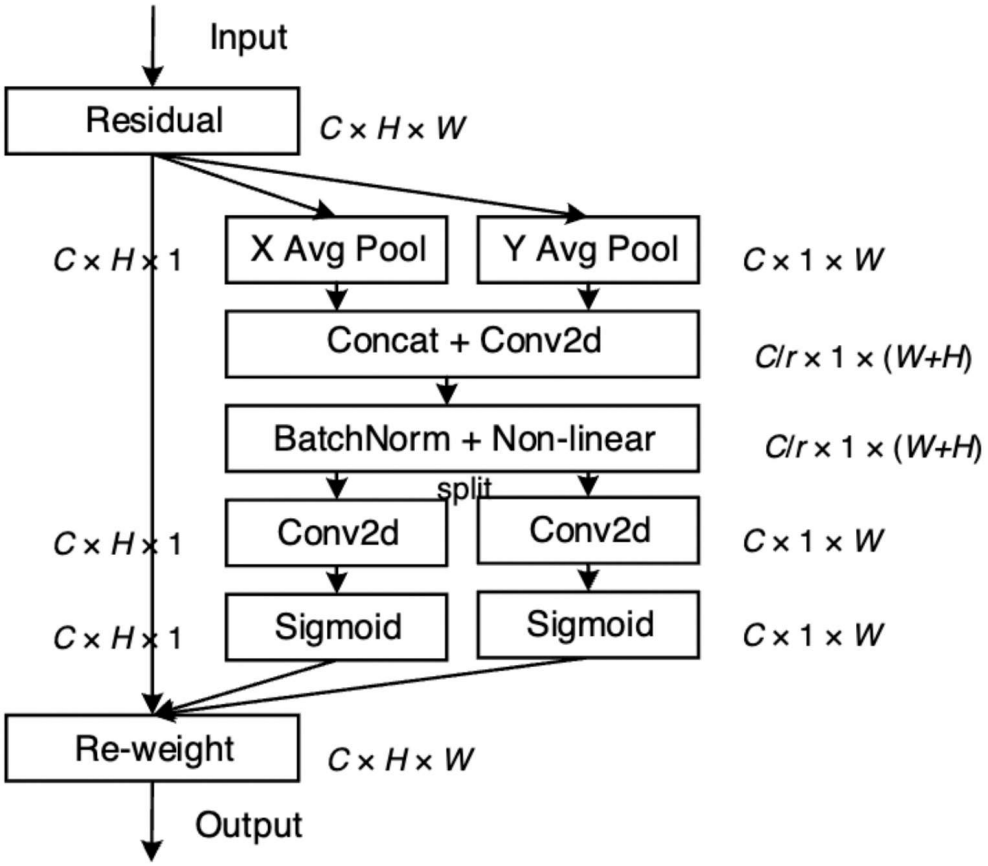


Figure 4. Coordinate attention module

Then, the two aggregated feature information will be concatenated, and through the 1×1 convolution transformation function F1 in Equation 3, an intermediate mapping of spatial information encoded in the vertical and horizontal directions will be obtained.

$$f = \delta\left(F_1([z^h, z^w])\right) \tag{3}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

Finally, the intermediate mapping will be split into two separate tensors f^h and f^w along the spatial dimension. They will be transformed into g^h and g^w with the same number of channels through two 1×1 convolution transformations in Equations 4 and 5, respectively. These two sums will be used as attention weights and ultimately fused in Equation 6 to obtain the output.

$$g^h = \sigma\left(F_h(f^h)\right) \tag{4}$$

$$g^w = \sigma\left(F_w(f^w)\right) \tag{5}$$

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{6}$$

Compared to traditional channel and spatial attention mechanisms, the coordinate attention has the following advantages. Firstly, it can capture inter-channel information and possesses direction-aware and position-sensitive capabilities, allowing for more accurate localization and identification of objects of interest. Secondly, the coordinate attention mechanism is flexible and lightweight, suitable for various classical mobile networks without imposing excessive computational burden on network inference speed. Lastly, the coordinate attention mechanism can significantly enhance the performance of downstream tasks in mobile networks, particularly in dense prediction and recognition tasks. Based on the Pascal VOC2007 dataset, the detection results of MobileNetV2 with different attention mechanisms are shown in Table 1.

Table 1. Object detection results on the Pascal VOC 2007 test set [18]

Backbone	Param(M)	M-Adds(B)	mAP(%)	Backbone
MobileNetV2	4.3	0.8	71.7	MobileNetV2
MobileNetV2 + SE	4.7	0.8	71.7	MobileNetV2 + SE
MobileNetV2 + CBAM	4.7	0.8	71.7	MobileNetV2 + CBAM
MobileNetV2 + CA	4.8	0.8	73.1	MobileNetV2 + CA

3.2 Weighted Feature Fusion

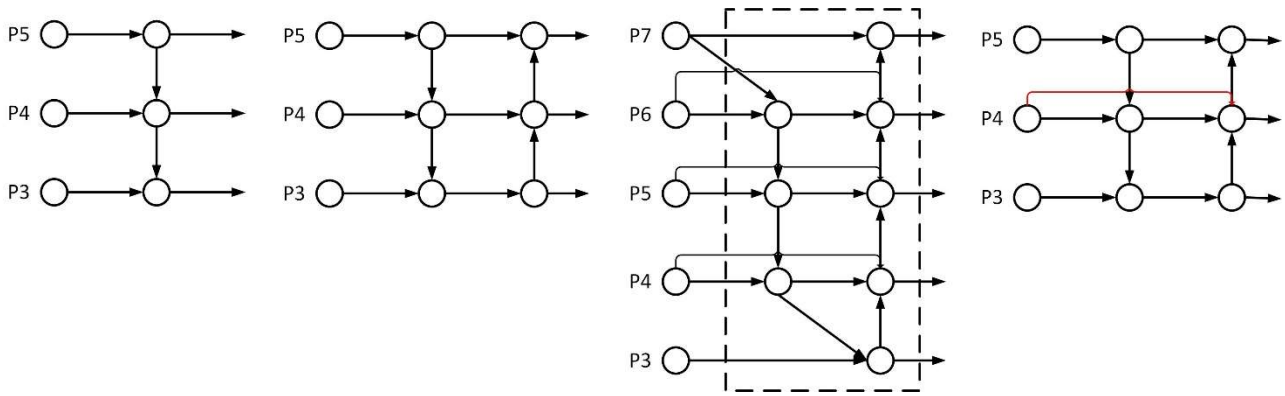
In object detection, further development of deep networks can enhance the model's capability to extract semantic features. However, as the network depth increases, the size of the output feature maps decreases, and the details of the image features are limited, which may pose a bottleneck for small target detection. Conversely, shallow networks can more effectively capture rich detailed information but lack sufficient semantic information, affecting the accuracy of the model in object classification.

To overcome these limitations, the academic community has begun to explore methods that combine deep and shallow features to extract more comprehensive and rich feature representations. Popular multi-scale feature fusion architectures include Feature Pyramid Networks (FPN) [20], Path Aggregation Network (PANet) [21], and Bi-directional Feature Pyramid Network (BiFPN) [22]. These architectural designs have shown great potential in improving detection performance.

As shown in Figure 5(a), the structure of the Feature Pyramid Network passes strong semantic information from top to bottom to fuse deep and shallow features. However, the FPN structure also has some deficiencies. It tends to focus on features of adjacent layers, and deep features may lose semantic information after multiple down sampling passes. To address this issue, the Path Aggregation Network structure, as shown in Figure 5(b), simultaneously employs two paths, top-down and bottom-up, for feature fusion. This structure shortens the information transmission path,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

effectively avoiding the loss of semantic information that may occur in FPN. However, this also increases computational overhead.



(a) FPN (b) PANet (c) BiFPN (d) WPANet

Figure 5. Structure diagram of multiscale feature fusion networks

In 2020, Tan proposed the Bi-directional Feature Pyramid Network based on Path Aggregation Network, the structure of which is shown in Figure 5(c), and brought the following improvements: Firstly, the nodes with only single input were removed as they did not contribute to feature fusion and only increased computational complexity. Secondly, skip connections were added, i.e., additional edges from the original input to the output nodes at the same level, in order to fuse more features without significantly increasing the cost. Thirdly, each bi-directional path formed by top-down and bottom-up was regarded as a network layer, and these network layers were repeated multiple times to achieve advanced feature fusion. Finally, different weights were allocated to each input based on their contribution to the final features, adjusting the proportion of different feature inputs in the output, and these weights were dynamically adjusted in real-time during network training.

The original YOLOv5s model used the FPN+PAN structure for neck network feature fusion, forming bi-directional paths from top-down and bottom-up. However, this approach is relatively simple, leaving further room for optimization in feature fusion. Inspired by the BiFPN, this section proposes to apply the design concept of BiFPN to construct the Weighted Feature Fusion Network (WPANet) in YOLOv5s.

Firstly, skip connections are added. When there are multiple fusion nodes in the feature fusion layer, multiple feature transmission edges are added between different nodes in the same layer. This allows the network to fuse more feature information with a small increase in parameters and computation, as shown in Figure 5(d) with the P4 annotated in red.

Secondly, weighted fusion. In the merging of multi-scale features with different resolutions, some previous object detection algorithms often treated the feature information outputted by each level equally, and adjusted their sizes to be consistent through up sampling or down sampling before adding them up. However, recent studies have found that the influence of input features with different resolutions on output features is not the same. In order to improve the algorithm's robustness and enhance the network's feature fusion capability, we introduced the idea of attention and introduced a hyperparameter for the feature information outputted by each level. This way, the network can learn the importance of features at different levels, thus optimizing the feature fusion process more effectively. For the traditional PANet, it aggregates multi-scale features simply by summing:

$$P_4^{\text{out}} = Conv\left(P_4^{\text{td}} + Resize(P_3^{\text{out}})\right) \tag{7}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

In the above formula, a common method used by Resize is to adjust the resolution of the feature map using up sampling or down sampling, so that the resolution of feature maps from different input scales matches. Conv is the convolution operation applied to the summed feature map to extract new features.

A weighted feature fusion method is used to integrate input feature maps from different resolutions. The corresponding weights for input layers with different resolutions are also different. By automatically learning the weight parameters of each input layer through the network, the overall feature information can be better represented. For the output parts of P4, rapid normalized fusion is performed separately, as shown in formula 8. Here, w_1' represents the weight of the current layer input, w_2' represents the weight of the output from the transition unit in the current layer, w_3' represents the weight of the output from the previous layer. ϵ is a hyperparameter used to prevent gradient disappearance. Conv represents the convolution operation on the overall calculation result.

$$P_4^{\text{out}} = Conv\left(\frac{w_1' \cdot P_4^{\text{in}} + w_2' \cdot P_4^{td} + w_3' \cdot Resize\left(P_3^{\text{out}}\right)}{w_1' + w_2' + w_3' + \epsilon}\right) \tag{8}$$

4. Dataset and Evaluations

4.1 Dataset

To validate the effectiveness of the algorithm in this article, the China Traffic Sign Detection Dataset 2021 (CCTSDB) [23] was selected. This dataset was created by Changsha University of Science and Technology for Comprehensive Transport Big Data, as a supplement to the original China Traffic Sign Dataset (CTSD), first published in 2017 and updated in 2021. The dataset consists of 16,356 images with resolutions ranging from 1000×350 to 1024×768. The images cover variations in factors such as scale, lighting, and noise, and are annotated with three major categories of signs: Mandatory, Prohibitory, and Warning. Mandatory signs are used to guide pedestrians and vehicles to travel in prescribed directions and locations; Prohibitory signs restrict or prohibit vehicles; Warning signs alert pedestrians and vehicles to dangerous areas, prompting them to slow down or proceed with caution. In the experiment, the dataset was randomly divided into a training set and a validation set in an 8:2 ratio.

Figure 6. Classification of traffic signs in China

4.2 Experiments Environment

The experiment in this article utilized the CCTSDB 2021 Traffic Sign Dataset, with image dimensions set at 640×640. The network parameters were configured as follows: the batch size (number of samples processed by the neural network in each training iteration) was set to 32, the momentum (weight decay parameter) was set to 0.98, the decay coefficient was set to 0.001, the training iterations

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

were set to 200, and the learning rate was set to 0.001. The experiment was conducted on a Linux server with specific configuration parameters for the training environment, as shown in Table 2.

Table 2. Experimental training environment configuration

OS	Ubuntu 18.04(Linux)
Deep learning framework	PyTorch1.7.0
CPU	15 vCPU AMD EPYC 7543 32-Core Processor
GPU	Nvidia GeForce RTX 3090(24GB)
Memory	80G
CUDA	11.5
Language	Python 3.8
Complier	Pycharm

4.3 Evaluation Metrics

Common methods for traffic sign detection and recognition are typically evaluated using mean Average Precision (mAP) and Frames Per Second (FPS) as the primary metrics. The mAP measures the average detection accuracy, while FPS indicates the processing speed of the model on image data, with higher FPS values indicating faster image processing by the model.Under different threshold conditions, mAP can have different definitions. In object detection tasks, the most common metrics are mAP@0.5 and mAP@[.5:.95], representing the average precision for all categories when the Intersection over Union (IoU) reaches 0.5 and ranges from 0.5 to 0.95, respectively. Additionally, this study also utilizes Precision and Recall as auxiliary evaluation metrics, with specific calculation formulas as follows.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$AP = \int_0^1 p(r)dr \tag{11}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{12}$$

TP, TN, FP, FN represent true positives (predicted positive and actually positive), true negatives (predicted negative and actually negative), false positives (predicted positive but actually negative), and false negatives (predicted negative but actually positive), respectively. In this study, m refers to the number of sample categories, with a value of 3.

5. Experiments and Analysis

5.1 Ablation Experiments

The experiment was conducted under the same environment settings as described in Section 4.2. A total of 200 epochs of training were performed on the CCTSDB 2021 traffic sign dataset to evaluate four different network models (YOLOv5s, YOLOv5s_C3CA, YOLOv5s_WPANet, YOLOv5s_C3CA+WPANet). The schematic diagram of the training loss function is shown in Figure

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

7. It can be observed from Figure 7 that in the initial stages of training, the loss function curves decrease rapidly. As the number of training iterations reaches approximately 50 epochs, the loss values of each network model gradually stabilize and eventually settle at around 0.024. It is noteworthy that the red loss function curve representing YOLOv5s_C3CA+WPANet consistently remains below the other curves throughout the training phase, indicating superior performance.

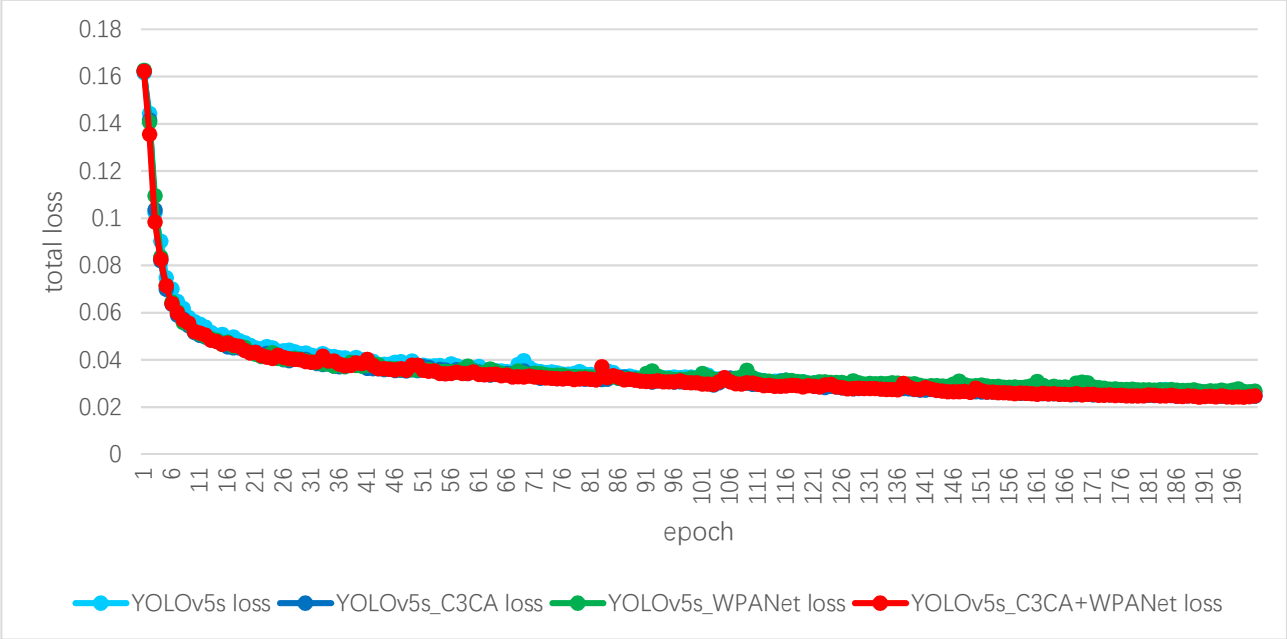


Figure 7. Illustration of Network Model Loss Function

The results of the ablation experiments are compared in Table 3. Based on the different network model parameters and test results in the table, the following conclusions can be drawn: Compared to YOLOv5s, YOLOv5s_C3CA, and YOLOv5s_WPANet models, the YOLOv5s_C3CA+WPANet model shows an increase in average precision mAP@0.5 by 3.83%, 2.41%, 1.08%, and an increase in average precision mAP@.5:.95 by 2.33%, 1.43%, 0.73%, respectively. Although there is a slight increase in model complexity and size, resulting in a decrease in detection speed to 98.04, it is sufficient to meet real-time requirements, demonstrating the effectiveness of the improved model.

Table 3. Model Parameters and Test Results Comparison Table

Network	mAP@.5(%)	mAP@.5:.95	FPS	FLOP(G)	Size(M)
YOLOv5s	74.90	47.66	126.58	16.5	13.7
YOLOv5s_C3CA	76.32	48.56	103.09	16.5	13.7
YOLOv5s_WPANet	77.65	49.26	123.46	16.7	13.8
YOLOv5s_C3CA+WPANet	78.73	49.99	98.04	16.7	13.8

To demonstrate the superiority of the innovative method proposed in this study, the YOLOv5s_C3CA+WPANet network algorithm in traffic sign detection and recognition tasks, a detailed performance comparison was conducted between this algorithm and YOLOv5s_C3CA, YOLOv5s_WPANet, and the original YOLOv5s model. Specifically, the precision (Precision) and recall (Recall) of these four models were evaluated on the CCTSDB 2021 traffic sign dataset. Furthermore, precision-recall (P-R) curve graphs were plotted with recall as the x-axis and precision as the y-axis, intuitively displaying the performance of each model in Figure 6. Through this curve

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

graph, a direct comparison of the detection effects of the original YOLOv5s model, CA-enhanced YOLOv5s model, WPANet-integrated YOLOv5s model, and the method proposed in this study on the same traffic sign dataset can be made.

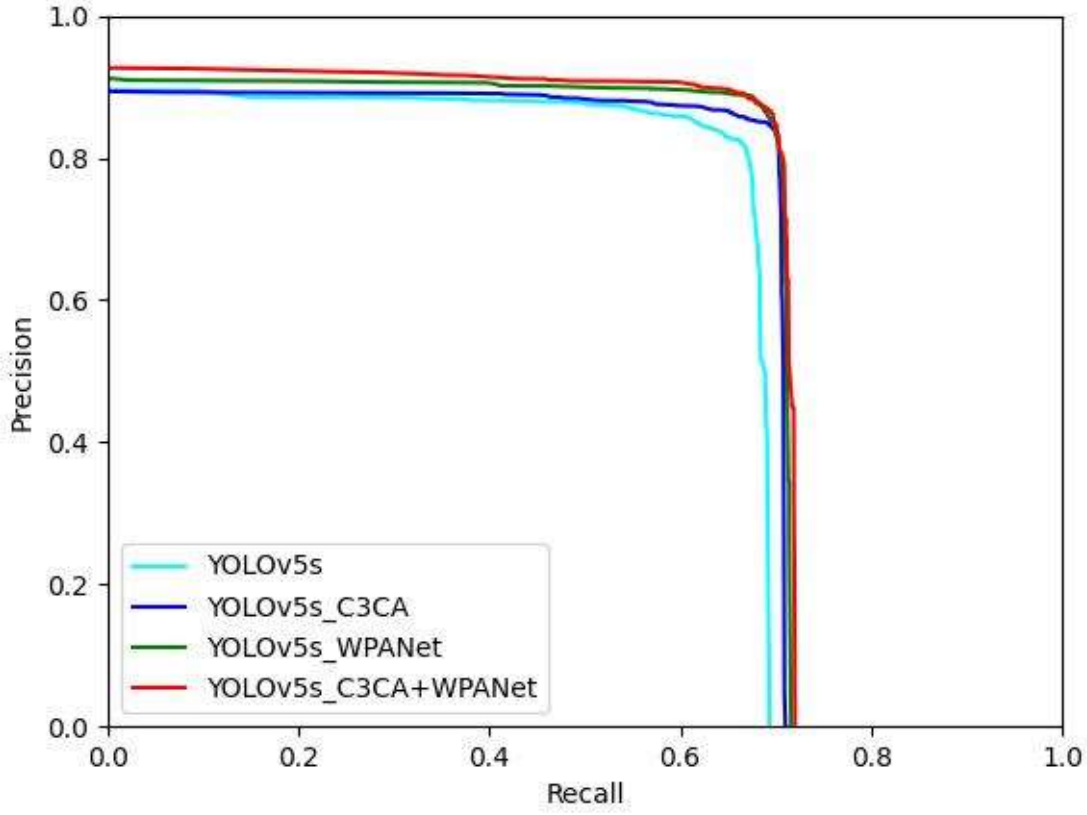


Figure 8. P-R Curve

It is evident from Figure 8 that the improved algorithm proposed in this chapter, YOLOv5s_C3CA+WPANet (red curve), not only almost completely covers the trajectory representing the YOLOv5s_WPANet detection model (green curve), but also consistently remains above the curve representing the YOLOv5s_C3CA model (blue curve), and clearly encompasses the cyan curve representing the original YOLOv5s model. This visual evidence demonstrates that the adopted improved algorithm has significant advantages in traffic sign detection and recognition tasks.

To further explore the detection effects of the improved network model on various categories, this study summarized and compared the results of different categories, as shown in Table 4. Compared to the original YOLOv5s network model, the improved network model shows significant improvements in Precision, Recall, mAP@0.5, and mAP@0.5:0.95 indicators. Taking the Mandatory category as an example, compared to YOLOv5s, YOLOv5s_C3CA+WPANet has improvements of 2.2%, 4.1%, 2.1%, and 1.1% in Precision, Recall, mAP@0.5, and mAP@0.5:0.95, respectively. In the Prohibitory category, these indicators have improved by 0.8%, 1.8%, 2.9%, and 1.8%, respectively. Particularly noteworthy is that in the Warning category, the improvements in these performance indicators for YOLOv5s_C3CA+WPANet are 7.8%, 3.7%, 6.4%, and 4.1%, respectively. The comprehensive data shows that the improved network model significantly enhances the detection effects of various traffic sign categories and performs excellently in precision and recall.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

Table 4. Table of Model Detection Results for Different Categories

Network	Categories	P/%	R/%	mAP@.5/%	mAP@.5:.95/%
	Mandatory	87.80	60.40	68.50	45.20
YOLOv5s	Prohibitory	91.90	66.10	75.30	48.90
	Warning	85.90	76.30	81.00	48.90
	Mandatory	90.00	64.50	70.60	46.30
YOLOv5s_C3CA+WPANet	Prohibitory	91.70	67.90	78.20	50.70
	Warning	93.70	80.00	87.40	53.00

5.2 Comprehensive Comparison Experiment

This study compares the performance of the optimized YOLOv5s_C3CA+WPANet network model with SSD, Faster R-CNN, and earlier versions of YOLO. The evaluation is based on various aspects, as indicated in the experimental results presented in Table 5. The optimized network model demonstrates significant advantages across multiple dimensions. While it may exhibit lower detection accuracy compared to the latest YOLOv8s version, it maintains comparable detection speed and notably outperforms in terms of precision. Overall, the proposed algorithm in this paper proves to be more precise and efficient in identifying traffic signs in the CCTSDB 2021 dataset, thereby contributing to the enhancement of traffic safety.

Table 5. Different Model Parameters and Test Results Comparison Table

Network	P/%	R/%	mAP@.5/%	FPS/(frame/s)
Faster R-CNN[23]	84.43	54.98	56.58	4.87
SSD[23]	86.47	27.74	49.2	22.33
YOLOv3[23]	84.63	42.71	50.48	20.34
YOLOv4[23]	76.16	52.50	51.69	16.55
YOLOv5s	88.60	67.60	74.90	126.58
YOLOv5s_C3CA+WPANet	91.80	70.80	78.70	98.04
YOLOv8s	87.90	74.50	82.30	106.38

5.3 Visualisation of Comparative Experiments

In this section, a subset of the CCTSDB dataset is selected for a comparative analysis of traffic sign recognition under various conditions, including normal sunny weather, nighttime rainy conditions, and heavy fog. The comparative results are depicted in Figure 9. Under normal sunny conditions, Figures 9(a) and 9(b) illustrate the recognition of mandatory and prohibitory traffic signs. The improved algorithm (YOLOv5s_C3CA+WPANet) shows a 3% increase in confidence when identifying prohibitory signs compared to the original algorithm (YOLOv5s). Additionally, the original algorithm exhibits a problem of missing mandatory signs, while the improved algorithm accurately recognizes such signs with a confidence level of 38%. In heavy fog conditions, as shown in Figures 9(c) and 9(d), both mandatory and prohibitory traffic signs are recognized. The improved algorithm demonstrates a 4% increase in confidence when identifying prohibitory signs compared to the original algorithm. Similarly, the original algorithm experiences issues with missing mandatory signs, whereas the improved algorithm achieves accurate recognition with a confidence level of 41%. In the case of nighttime rainy conditions, as depicted in Figures 9(e) and 9(f), prohibitory traffic signs are present. The improved algorithm shows a 3% increase in confidence when identifying prohibitory

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

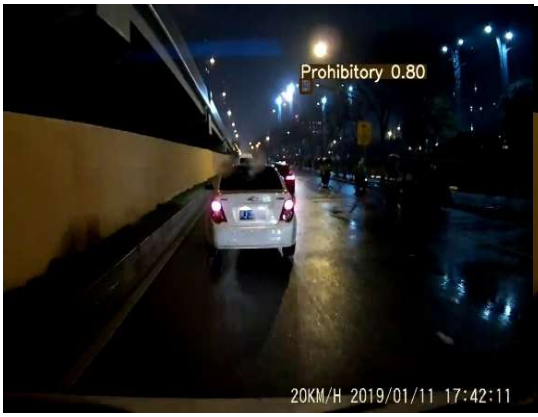
signs compared to the original algorithm. These results indicate that even in challenging and complex environments, the improved algorithm maintains high accuracy and exhibits robustness.

(a) Sunny Weather of Original Algorithm

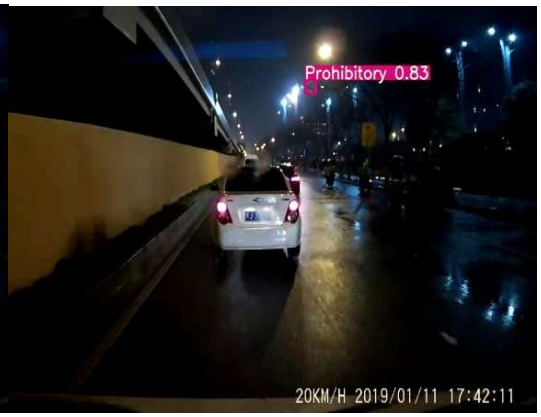
(b) Sunny Weather of Improved Algorithm

(c) Heavy fog of Original Algorithm

(d) Heavy fog of Improved Algorithm



(e) Nighttime Rainy of Original Algorithm



(f) Nighttime Rainy of Improved Algorithm

Figure 9. Comparison diagrams of different environmental detection effects

6. Conclusion

The existing challenges in traffic sign detection and recognition, such as small targets and diverse sizes, were addressed in this study. Building upon YOLOv5s, the following improvements were made: the use of the feature enhancement module C3CA to improve the focusing ability on key areas, and the adoption of a weighted feature fusion network to enhance the fusion capability of multi-size features. The optimized algorithm has enhanced the accuracy of traffic sign detection. Through ablation experiments and comparative analysis, it has been demonstrated that the improved algorithm has achieved a certain level of improvement in detection accuracy and can achieve real-time detection speed. Future research could focus on applying the optimized algorithm to traffic sign detection and recognition in embedded systems engineering, and further exploration and application.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0010

References

[1] Qian R , Zhang B , Yue Y ,et al.Traffic sign detection by template matching based on multi-level chain code histogram[J].International Conference on Fuzzy Systems and Knowledge Discovery, 12th International Conference on Fuzzy Systems and Knowledge Discovery(FSKD), 2015.

[2] He X , Dai B .A new traffic signs classification approach based on local and global features extraction[C]//International Conference on Information Communication & Management.IEEE, 2016:121-125.

[3] Zou Z , Shi Z , Guo Y ,et al.Object Detection in 20 Years: A Survey[J]. 2023.

[4] Girshick R , Donahue J , Darrell T ,et al.Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J].IEEE Computer Society, 2014.

[5] Girshick R .Fast R-CNN[C]//International Conference on Computer Vision.IEEE Computer Society, 2015.

[6] Ren S , He K , Girshick R ,et al.Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[7] Redmon J , Divvala S , Girshick R ,et al.You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision & Pattern Recognition.IEEE, 2016.

[8] Redmon J , Farhadi A .YOLO9000: Better, Faster, Stronger[C]//IEEE Conference on Computer Vision & Pattern Recognition.IEEE, 2017:6517-6525.

[9] Redmon J , Farhadi A .YOLOv3: An Incremental Improvement[J].2018.

[10] Bochkovskiy A , Wang C Y , Liao H Y M .YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.

[11] Wei L , Dragomir A , Dumitru E ,et al.SSD: Single Shot MultiBox Detector[J].Springer, Cham, 2016.

[12] Simonyan K , Zisserman A .Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[13] Yang T , Long X , Sangaiah A K ,et al.Deep detection network for real-life traffic sign in vehicular networks[J].Computer Networks, 2018, 13(6):95-104.

[14] Jin Y , Fu Y , Wang W ,et al.Multi-Feature Fusion and Enhancement Single Shot Detector for Traffic Sign Recognition[J].IEEE Access, 2020, 8:38931-38940.

[15] Zhang Z , Wang H , Zhang J ,et al.A vehicle real-time detection algorithm based on YOLOv2 framework[J].2018:1-6.

[16] Sichkar V N , Kolyubin S A .Real time detection and classification of traffic signs based on YOLO version 3 algorithm[J].Scientific and Technical Journal of Information Technologies Mechanics and Optics, 2020, 20(3):418-424.

[17] Liu X, Jiang XK, Hu HC, et al. Traffic sign recognition algorithm based on improved YOLOv5s[J]. Proceedings of the 2021 International Conference on Control, Automation and Information Sciences (ICCAIS). Xi'an: IEEE, 2021:980–985.

[18] Hou Q , Zhou D , Feng J .Coordinate Attention for Efficient Mobile Network Design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:13713-13722.

[19] Tian Z , Shen C , Chen H ,et al. Focus: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision.2019:9627-9636.

[20] Lin T Y ,Dollár, Piotr, Girshick R ,et al.Feature Pyramid Networks for Object Detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2017: 2117-2125.

[21] Liu S , Qi L , Qin H ,et al.Path Aggregation Network for Instance Segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2018: 8759-8768.

[22] Tan M , Pang R , Le Q V .EfficientDet: Scalable and Efficient Object Detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).IEEE, 2020:10781-10790.

[23] Zhang J , Zou X , Li D K , et al.Jin Wang, R. Simon Sherratt, Xiaofeng Yu. CCTSDB 2021: A more comprehensive traffic sign detection benchmark. Human-centric Computing and Information Sciences, 2022, vol. 12, Article number:23.