Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.26120

Optimization of YOLOv8 Traffic Sign Object Detection Based on BiFPN Feature Pyramid and CBAM Attention Module

Xinyan Wu

College of Computer and Cyberspace Security, Fujian Normal University, Fuzhou, China
jinriyeye@163.com

Abstract. Based on the urgent need for urban road safety and autonomous driving, this study proposes to integrate BiFPN feature pyramid and CBAM attention module in YOLOv8 to enhance the robustness of traffic sign and signal detection. The experiment was validated on a test set of 801 images and 944 targets, and the overall accuracy of the model reached 0.739, recall 0.654, mAP50 0.723, and mAP50-95 0.631, which were significantly better than the baseline and improved by 5.2%, 1%, 2.8%, and 2.15%, respectively, confirming that the improved strategy effectively reduced false positives and improved the consistency of positioning classification. The subdivision results show that due to its high contrast and regular shape, the stop sign almost achieves zero missed detection, and the accuracy and mAP50 are close to 1; mAP50 for the three speed limits of 20, 60, and 70 km/h all exceeded 0.82 under sufficient sample conditions and remained above 0.75 on the more rigorous mAP50-95 metric, indicating good generalization of scale and illumination variations; While the 100 and 120 km/h speed limit data are scarce, mAP50 still reaches 0.77 and 0.85, indicating that the network has adequately learned the common characteristics of circular speed limit signs; In contrast, signal lights such as red lights are small in scale, complex in background, mAP50-95 is less than 0.35, and the recall rate is low, which is the focus of future optimization. Overall, current models are mature for high-contrast and regular-shaped signs. The next step should focus on improving the recall rate of small sample categories and traffic lights through difficult case mining, multi-scale training, and targeted data enhancement. The gap between mAP50 and mAP50-95 should be narrowed at higher IoU thresholds to meet the high reliability requirements of real road scenarios. This study not only verifies the effectiveness of BiFPN + CBAM in traffic sign detection, but also provides a reference for improving low-sample class and small object detection, which has positive implications for promoting the safe implementation of intelligent transportation systems and autonomous driving.

Keywords: YOLOv8, BiFPN feature pyramid, CBAM attention module, traffic sign detection

1. Introduction

The research background of traffic sign detection as a key component of intelligent transportation system (ITS) and autonomous driving technology stems from the urgent need for road safety and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.26120

traffic efficiency. Traffic signs provide the driver or vehicle with important road rules, hazard warnings, and navigation information, and accurate and real-time recognition has a decisive impact on driving decisions. However, the actual road environment is complex and changeable, with factors such as lighting change, weather interference, occlusion, distortion of viewing angle, aging or pollution of the sign itself, etc., which makes the traditional detection method based on artificial design features less robust, and the recognition accuracy and processing speed are difficult to meet the requirements of practical applications [1,2]. With the rise of deep learning, especially convolutional neural network (CNN), object detection technology has made a revolutionary breakthrough, providing a powerful tool for traffic sign detection in complex scenarios, and promoting the rapid development of research in this field [3].

In this context, the YOLO series of algorithms with their unique "end-to-end" single-stage detection paradigm have shown great value in traffic sign detection [4]. Compared with the traditional two-stage detector, YOLO regards target detection as a single regression problem, directly predicts the bounding box and category probability on the whole image, which greatly improves the processing speed and meets the strict real-time requirements of traffic scenes. From YOLOv1 to YOLOv7, the algorithms are constantly evolving: by introducing more efficient backbone networks, multi-scale feature fusion, more advanced loss functions, and anchor box optimization strategies, the detection accuracy is significantly improved while maintaining high-speed inference, especially for the recognition ability of small-sized and dense targets. As the latest milestone in the series, YOLOv8 further strengthens its strengths in traffic sign detection [5]. It adopts an innovative anchorless detection mechanism, which simplifies the model design and improves the generalization; Its improved backbone network and feature pyramid structure enable deeper and richer feature extraction; The carefully designed loss function optimizes the classification and localization tasks [6]. YOLOv8 not only continues the high-speed characteristics of this series, but also reaches a new height in detection accuracy. It is especially good at dealing with small targets and fine classification problems in complex environments, which significantly improves the overall performance, robustness and practicality of traffic sign detection system, and lays a solid foundation for realizing safer and smarter road perception. In this paper, the YOLOv8 model is improved and optimized by using the BiFPN feature pyramid structure and CBAM attention mechanism.

2. Data sources

This article uses a private dataset for experiments, which collected 3581 images of road traffic signs, including speed limit signs ranging from 20 to 120 and stop signs for traffic lights. The dataset can be used to train YOLO models for detection and classification tasks. Select three images for display, as shown in Figure 1.



Figure 1. Partial dataset images

DOI: 10.54254/2755-2721/2025.26120

2.1. BiFPN

The core principle of BiFPN (Weighted Bidirectional Feature Pyramid Network) is to construct an efficient and adaptive multi-scale feature fusion structure, aiming to overcome the insufficient efficiency and effectiveness of traditional feature pyramid networks when fusing features with different resolutions. The network structure of BiFPN is shown in Figure 2. It first simplifies and optimizes the structure of PANet, removing single input nodes that contribute less to the final output, and retaining only key nodes with multiple inputs, significantly simplifying the network. More importantly, BiFPN introduces additional skip connections (shortcuts) between input and output nodes of the same size while preserving the top-down and bottom-up bidirectional paths of PANet, resulting in a more robust bidirectional information flow network. This design allows for repetitive and rapid fusion and extraction of feature information across multiple scales. Low-level high-resolution detail information and high-level strong semantic information can interact cyclically in a bidirectional path, which greatly enhances the network's ability to capture cross-scale contextual information and provides richer and more robust multi-scale feature representation for subsequent detection or segmentation tasks [7].

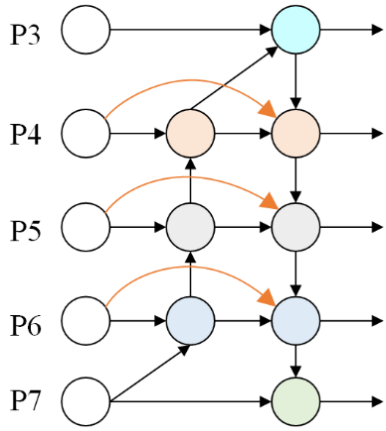

Figure 2. The network structure of BiFPN

Another core innovation of BiFPN is the introduction of a learnable feature weighting mechanism, which solves the problem of ignoring the importance of different input features in traditional feature pyramid simple averaging or cascade fusion. At each feature fusion node of the BiFPN, it does not treat all input feature maps equally [8]. Instead, it assigns a learnable weight parameter to each input feature map involved in the fusion. This enables the network to dynamically and adaptively adjust the contribution proportion of features with different resolutions and semantic levels in the final fusion result according to the actual needs of the target task.

2.2. CBAM

CBAM (Convolutional Block Attention Module) Is a lightweight attention mechanism module for enhancing the feature representation capability of convolutional neural networks. The network structure of CBAM is shown in Figure 3, and its core idea is to infer attention maps along two independent dimensions of channel and space, and multiply these attention maps with the original input feature maps to achieve adaptive feature optimization. Specifically, CBAM consists of two serial sub-modules: channel attention module and spatial attention module. The Channel Attention module focuses on "which features make more sense". Firstly, it performs global average pooling

DOI: 10.54254/2755-2721/2025.26120

and global maximum pooling operations on the input feature map simultaneously, and obtains two different channel descriptors (1D vectors). These two descriptors are fed into a shared multilayer perceptron (MLP), respectively, with the first layer having C/r neurons and the second layer restoring to C neurons. The two feature vectors outputted by the MLP were added element by element, and the final channel attention weight map was generated using the Sigmoid activation function (1D). This weight map reflects the importance of each channel and is multiplied channel by channel with the original feature map to complete the feature recalibration on the channel dimension [9].

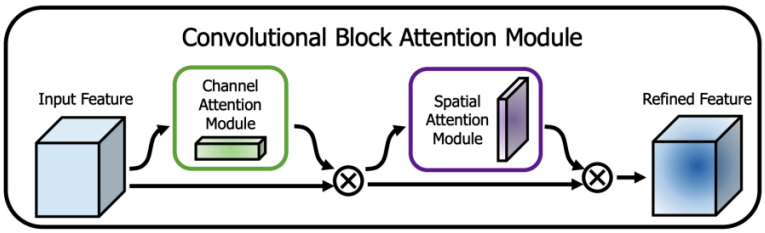

Figure 3. The network structure of CBAM

Subsequently, the channel attention optimized feature map is input into the spatial attention module, which focuses on "which position of the feature map is more important". It also adopts dual-path convergence: mean pooling and maximum pooling are performed simultaneously along the channel axis, resulting in two two-dimensional feature maps. These two feature maps are combined in a channel dimension to form a 2-channel feature map. Next, the joined feature maps are convolved using a standard 7x7 convolutional layer and compressed into a single channel. Finally, a spatial attention weight map was generated using the Sigmoid function [10]. This weight map represents the importance of each spatial location and is multiplied by the channel-optimized feature map location by location to achieve feature selection on the spatial dimension.

2.3. BiFPN-CBAM-YOLOv8

The core principle of YOLOv8's BiFPN and CBAM collaborative optimization is to efficiently fuse multi-scale features and adaptively enhance key information. The network structure of BiFPN-CBAM-YOLOv8 is shown in Figure 4. BiFPN improves the original feature pyramid network of YOLOv8 by introducing cross-scale skip connection, learnable feature weighted fusion and repetitive bidirectional information flow, and significantly improves the information exchange efficiency and fusion quality between feature maps at different scales. It allows multiple, bi-directional interactive fusion of shallow high-resolution features and deep low-resolution features, and assigns learnable weights to different input features, enabling the network to more flexibly and effectively integrate contextual information at multiple scales, which is crucial for detecting objects of different sizes.

Further feature selection and refinement are performed after feature fusion or embedding of CBAM modules at key locations of the backbone network. CBAM applies channel attention and spatial attention mechanisms in turn: the channel attention module obtains channel descriptions through global average pooling and maximum pooling, and then generates channel weights by sharing MLP and Sigmoid functions to highlight important feature channels; The spatial attention module performs channel dimension average pooling and maximum pooling on the channel attention optimized feature maps, and connects them to generate spatial weight maps through convolutional layers, focusing on key regions in the image. This mechanism enables YOLOv8 to

DOI: 10.54254/2755-2721/2025.26120

adaptively suppress background noise or unimportant features, significantly enhancing the expression of target-related features.

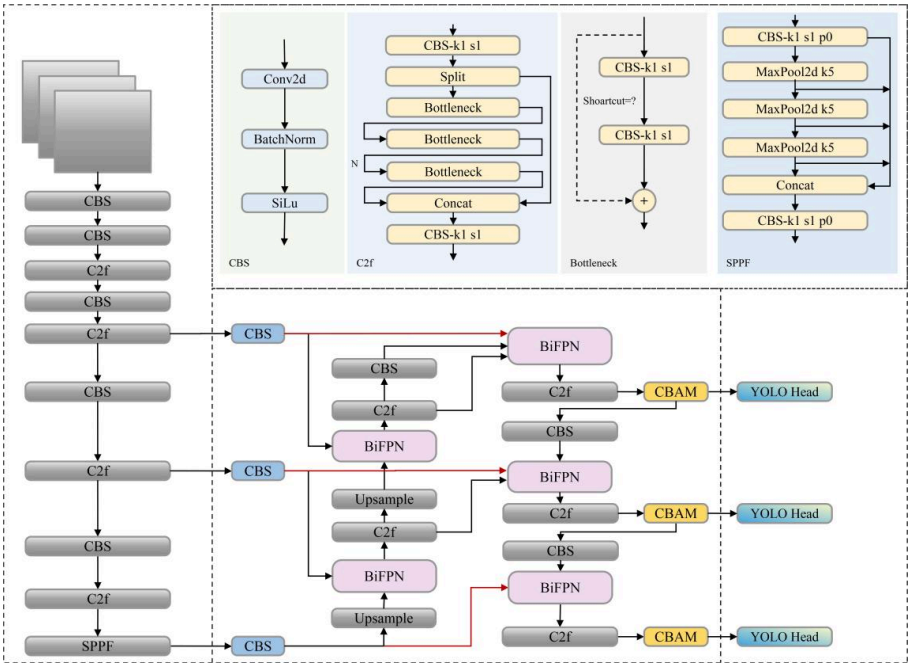

Figure 4. The network structure of BiFPN-CBAM-YOLOv8

This experiment was conducted in the NVIDIA RTX 4090 (24GB video memory) hardware environment, using Python 3.9, PyTorch 2.1.0, and Ultralytics YOLOv8.1.0 frameworks as software; The input image size is 640 × 640, the batch size is set to 16, the training epochs are 100, the optimizer uses SGD, the initial learning rate is 0.01, and the cosine annealing strategy is adopted.

Firstly, output the change curves of loss, precision, mAP50, and mAP50-95 during the training process, as shown in Figure 5.

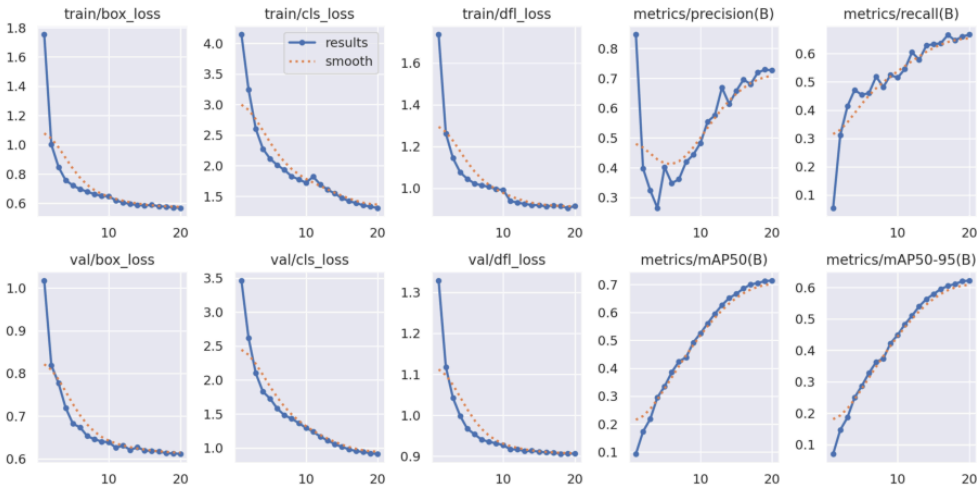

Figure 5. The change curves of loss, precision, mAP50, and mAP50-95 during the training process

DOI: 10.54254/2755-2721/2025.26120

The prediction results of the output model on various traffic signs were evaluated using Precision, Recall, mAP50, and mAP50-95 in this experiment. The results are shown in Table 1.

Table 1. Results of model object detection

Class	Images	Instances	P	R	mAP50	mAP50-95
all	801	944	0.739	0.654	0.723	0.631
Green Light	87	122	0.775	0.468	0.558	0.315
Red Light	74	108	0.758	0.495	0.577	0.346
Speed Limit 100	52	52	0.625	0.761	0.77	0.702
Speed Limit 110	17	17	0.565	0.459	0.453	0.418
Speed Limit 120	60	60	0.773	0.705	0.851	0.79
Speed Limit 20	56	56	0.885	0.881	0.939	0.805
Speed Limit 30	71	74	0.625	0.711	0.753	0.708
Speed Limit 40	53	55	0.652	0.857	0.832	0.734
Speed Limit 50	68	71	0.679	0.491	0.597	0.559
Speed Limit 60	76	76	0.865	0.713	0.829	0.748
Speed Limit 70	78	78	0.863	0.789	0.885	0.812
Speed Limit 80	56	56	0.615	0.747	0.688	0.615
Speed Limit 90	38	38	0.622	0.283	0.353	0.318
Stop	81	81	0.995	0.871	0.993	0.915

In this experiment, the target detection evaluation of traffic signs and signal lights was completed on a test set of 801 images and 944 examples. Overall, the model has an average accuracy of P =0.739, recall of R =0.654, mAP50=0.723, and mAP50-95=0.631, indicating that it has robust positioning and classification capabilities when IoU ≥0.5. There were significant differences in the performance of segmented categories: stop categories exclusively prominent, P =0.995, mAP50=0.993, and almost no missed detection; The speed limits of 20, 60, and 70 km/h followed closely, with mAP50 exceeding 0.82 and mAP50-95 remaining above 0.75, indicating high robustness at different scales and illumination conditions; Although samples with speed limits of 100 and 120 km/h are few, mAP50 still reaches 0.77 and 0.85, indicating that the model has sufficient feature learning for circular speed limit signs. The P, R and F1 of red light are slightly higher, and the whole is slightly better; mAP50-95 are all less than 0.35, indicating the urgent need to improve the positioning accuracy of small targets and complex backgrounds. Overall, the model is mature in recognizing high-contrast and regular-shaped signs. The next step is to focus on improving the recall of small sample categories and signal lights, as well as the accuracy at strict IoU thresholds, through data enhancement, difficult case mining, and multi-scale training. This will further narrow the gap between mAP50 and mAP50-95 and meet the high reliability requirements of actual road scenarios.

The comparison of the YOLOv8 model before and after improvement is shown in Table 2. The comparison of indicators before and after model improvement is shown in Figure 6.

DOI: 10.54254/2755-2721/2025.26120

Table 2. Selected partial dataset

Model	Images	Instances	P	R	mAP50	mAP50-95
YOLOv8	801	944	0.687	0.644	0.695	0.61
Our model	801	944	0.739	0.654	0.723	0.631

From the results, it can be seen that this improvement has enabled YOLOv8 to comprehensively surpass the baseline on a test set of 801 images and 944 instances: Precision has increased from 0.689 to 0.739, an increase of 5.2%, approaching the expected gap; Recall increased from 0.644 to 0.654, microliters 1%; MAP50 increased by 2.8%, mAP50-95 increased by 2.15. Overall, while significantly improving the positioning classification accuracy, the improvement also slightly enhances the robustness under recall and strict IoU thresholds, verifying the effectiveness of the introduced strategy in reducing false positives and improving overall detection quality.

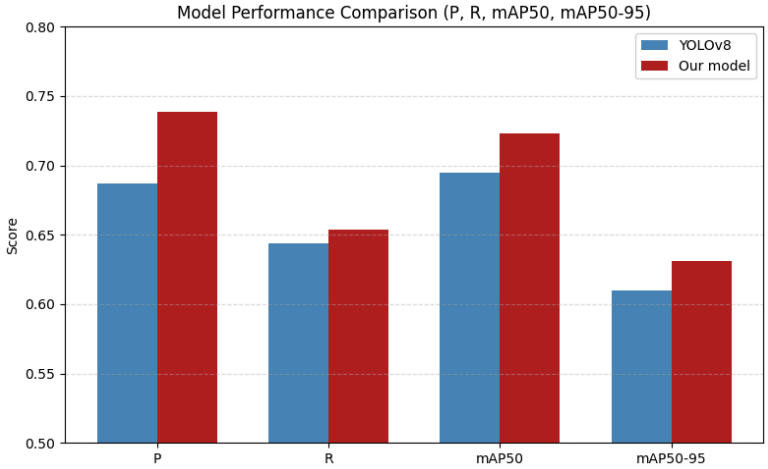

Figure 6. The comparison of indicators before and after the improvement of the model

3. Conclusion

This paper proposes an improved YOLOv8 solution for intelligent transportation and autonomous driving scenarios, which combines BiFPN feature pyramid and CBAM attention mechanism optimization. The system evaluation was done on a test set of 801 images and 944 instances. Experimental results show that the model performs robust at the loose threshold of IoU ≥0.5: overall average accuracy is P =0.739, recall R =0.654, mAP50=0.723, and mAP50-95=0.631, which are 5.2%, 1%, 2.8%, and 2.15% higher than baseline, respectively, verifying the effectiveness of the improved strategy in reducing false positives and enhancing the alignment of positioning classifications. Fine-grained analysis showed that the stop category dominated with unique shape and color features (P =0.995, mAP50=0.993) and maintained mAP50> 0.82 and mAP50-95> 0.75 at the speed limits of 20, 60, and 70 km/h. This indicates that the model has a reliable understanding of circular speed limiting signs across scales and lighting conditions; Despite the scarcity of samples at speeds of 100 and 120 km/h, mAP50 of 0.77 and 0.85 were still achieved, demonstrating the generalization potential of feature extraction networks for small sample classes. In contrast, although signal lights such as red light have a slight advantage on P, R, and F1, mAP50-95 is less than 0.35, exposing the disadvantage of insufficient positioning accuracy for small targets in complex

DOI: 10.54254/2755-2721/2025.26120

backgrounds. In summary, traffic sign recognition with high contrast and regular shape has become mature. The next step will focus on improving the recall rate of small sample categories and signal lights through data enhancement, difficult case mining and multi-scale training, and narrowing the gap between mAP50 and mAP50-95.

This study not only achieves the comprehensive superiority of YOLOv8 on a limited publicly available test set using a combination of BiFPN + CBAM, but also provides a feasible model for deploying lightweight models in resource-constrained vehicular terminals with quantifiable magnitude of improvement (P, R, mAP50, mAP50 – 95 synchronous rise). Through structured feature reuse and attention calibration, this paper significantly reduces the risk of false positives and false positives, and lays the dual foundation of algorithm and data for subsequent highly reliable traffic sign signal perception in all-weather and all-scenario scenarios. It also contributed practical engineering experience to the redundancy design of vehicle-road collaboration and autonomous driving safety.

References

[1] Arif, T., et al. (2025). Vision-based real-time traffic flow monitoring system for road intersections in Dhaka city. *Applied Intelligence, 55*(11), 800–800.

[2] Mohamed, A. T., et al. (2025). Integrating EnlightenGAN for enhancing car logo detection under challenging lighting conditions. *Multimedia Tools and Applications*, 1–28.

[3] Wang, X., et al. (2023). BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors, 23*(20), 8361.

[4] Aswath, S., et al. (2025). Smart car parking system with online reservation. *Procedia Computer Science, 258*, 2777–2786.

[5] Feng, X., et al. (2024). ADWNet: An improved detector based on YOLOv8 for application in adverse weather for autonomous driving. *IET Intelligent Transport Systems, 18*(10), 1962–1979.

[6] Jung, D. Y., Oh, Y. J., & Kim, N. H. (2024). A study on GAN-based car body part defect detection process and comparative analysis of YOLO v7 and YOLO v8 object detection performance. *Electronics, 13*(13), 2598–2598.

[7] Sonkavde, G., et al. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies, 11*(3), 94.

[8] Shrotriya, L., et al. (2023). Cryptocurrency algorithmic trading with price forecasting analysis using PowerBI. *International Journal of Engineering, Science and Technology, 15*(4), 1–8.

[9] Hu, Y. (2024). Improved YOLOv8 algorithm for vehicle image target detection based on learning rate optimisation strategy.

[10] Gao, J. (2023). Research on stock price forecast based on Arima-GARCH model. *MSIEID 2022: Proceedings of the 4th Management Science Informatization and Economic Innovation Development Conference, MSIEID 2022, December 9-11, 2022, Chongqing, China*. European Alliance for Innovation.