

# Research on the Improved YOLOv8 Algorithm for Detecting Multi - target Road Obstacles

Jiangfeng Sun\*, Pengfei Feng <sup>a</sup>, Yanyang Zeng <sup>b</sup>

Henan Polytechnic University, College of Computer Science and Technology, Jiaozuo 454000, China

\* Corresponding author: Jiangfeng Sun (Email: sunjiangfeng@hpu.edu.cn), <sup>a</sup>956361280@qq.com, <sup>b</sup>zengyy@hpu.edu.cn

**Abstract:** Different road obstacles may have diverse appearance, size, quantity, and location characteristics, and they are often accompanied by occlusion and overlapping phenomena. These factors result in the current object detection algorithms performing poorly in terms of detection accuracy and robustness when dealing with multiple objects and small objects with indistinct features. To address this issue, a lightweight road obstacle detection algorithm based on the improved YOLOv8n is proposed. Firstly, the MFObstacle road obstacle object detection dataset is created, which expands the types and quantities of obstacles for detection. Secondly, in the Neck part of the network, the GLSA (Global-to-Local Spatial Aggregation) module is added to enhance the multi-scale spatial aggregation mechanism of the network and improve the model's sensitivity and recognition ability for objects of different scales. Then, the BiFPN (Bidirectional Feature Pyramid Network) is integrated. BiFPN can dynamically adjust the weights according to the input data, making the network structure simpler and the computational efficiency higher. Finally, MPDIoU is used. MPDIoU takes into account the overlapping or non-overlapping regions, the distance between the center points, as well as the deviations in width and height, and simplifies the calculation process. Compared with the original YOLOv8n algorithm, the mAP0.5 index of this algorithm is increased by 2.2% to reach 85.7%; the number of network parameters is reduced by 30.4% to 4.6MB; and the computational complexity is reduced by 6% to 7.6GFLOPs. This algorithm improves the detection accuracy while reducing the number of parameters and computational complexity, and can be better applied to the task of multi-object road obstacle detection.

**Keywords:** Multi-object Obstacle Detection; YOLOv8; Improved C2f Module; GLSA; BiFPN.

## 1. Introduction

In the past few years, along with the swift advancement of artificial intelligence technology, the automotive sector is accelerating the pace of its intelligent transformation [1]. However, achieving autonomous driving still faces many technical challenges, among which environmental perception is crucial for ensuring driving safety and enhancing the driving experience [2]. The complexity of the road environment and the uncertainties brought about by natural factors make accurate and rapid detection of road obstacles a key factor for autonomous driving systems to make correct decisions and ensure safe driving. At present, object detection algorithms are predominantly categorized into two types. The first type is two - stage detection algorithms, with Fast RCNN [3] and Faster R - CNN [4] being their typical representatives. These algorithms boast high detection precision; however, they operate at a relatively slow speed, rendering them inappropriate for obstacle recognition tasks that demand real - time detection. The second type consists of one - stage detection algorithms, of which YOLO and SSD are the most well - known examples. These algorithms do not need to generate candidate regions and can output results in one step, so they perform well in terms of speed and also have quite high detection accuracy. Huu et al. used a pre-trained YOLO model to identify road obstacles, and the mean average precision (mAP) of obstacle detection reached 74.1%, but the number of detectable road obstacle categories was relatively small[5]. In order to enable intelligent vehicles to accurately detect pedestrians in complex scenarios, Cao et al. made many improvements to the YOLOv3 network, including adjusting the network cell size and adopting the k-means clustering algorithm, etc. They solved the problems that the

original YOLOv3 could not fully recognize the scale features of objects and had inaccurate object localization, and the mAP value for detecting pedestrians reached 90.42% [6]. Shivayogi et al. deployed the YOLOv4 model architecture on the NVIDIA Jetson Nano to recognize traffic signs on the road in real time. The accuracy was between 54.68% and 76.55%, and the speed could reach 32.5FPS (frames per second). However, for small-sized traffic signs, there were still problems of missed detection and false detection [7]. To implement vehicle monitoring and detection in videos on video surveillance devices, Wang and his colleagues substituted the backbone feature extraction network of YOLOv5 with MobileNetV2. This approach addressed the issue of the relatively large number of parameters in the original YOLOv5 network and decreased the size of the network model to 0.9MB [8]. Confronted with the challenge that current pavement defect detection algorithms are unable to achieve a simultaneous balance between detection speed and accuracy, Huang et al. made enhancements to the network architecture of the YOLOv7 algorithm, successfully solving the problems of slow speed and high computational load of the original YOLOv7 model, achieving an FPS of 80 frames per second and an mAP of 91% [9].

In summary, the YOLO series, including the latest YOLOv8, is widely used in road image processing. But when detecting multi - object road obstacles, YOLOv8[10] still has the same issues as other YOLO algorithms. Thus, this study improves YOLOv8 for real - time and accurate obstacle detection in driving environments.

## 2. Establishment of the Multi-object Road Obstacle Detection Network

### 2.1. Principle of the YOLOv8 Algorithm

In 2023, Ultralytics released the YOLOv8 model with n/s/m/l/x versions for different scenarios. Its structure comprises Input, Backbone, Neck, and Head (Fig. 1). At the input, Mosaic augmentation and other techniques boost robustness. The C2f[11] in the backbone enables rich feature learning but has high complexity. The PAN - structured neck fuses features via sampling and stitching, while the head decouples classification and detection, enhancing bounding box prediction accuracy

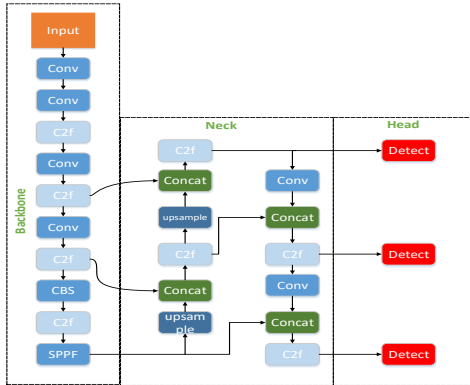


Figure 1. Structure of Fog-Mod-YOLOv8n

### 2.2. Multi-object Road Obstacle Detection Algorithm

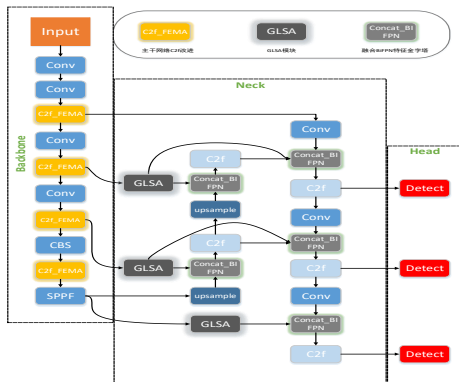
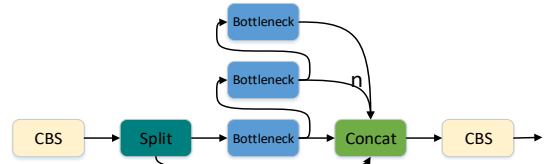


Figure 2. The Network Structure Diagram of the Improved YOLOv8n

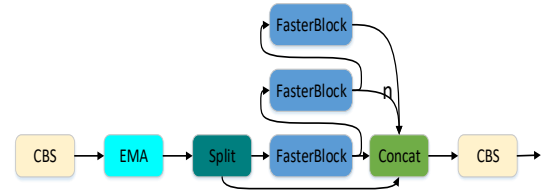
### 2.3. Lightweight Design of the Backbone Network

The YOLOv8 network utilizes the concept of multi-scale feature fusion, which is advantageous for making more thorough use of features at various scales. Nevertheless,

because of the increase and combination of the number of channels, there is a possibility that multiple channels might carry similar or the same information, thus giving rise to the problem of feature map redundancy. In order to tackle this problem and improve the network's detection speed while ensuring the detection accuracy, the C2f\_FEMA[12] is devised. It incorporates the EMA attention mechanism and substitutes the Bottleneck structure with FasterBlock[13]. By adopting this method, the weight of the model is reduced, the computational efficiency is enhanced, and the accuracy is still maintained. The module made up of C2f\_FEMA is employed to replace the C2f module within the backbone network. The structure of the module consisting of C2f\_FEMA is illustrated in Figure 3(b).



(a)The Structure Diagram of the C2f Module



(b)The Structure Diagram of the C2f\_FEMA Module

Figure 3. Improvement of the C2f Module

### 2.4. Introduction of the GLSA Module

The GLSA [14] network structure adapts to detection tasks of different scales and complex backgrounds by combining global and local features. It aggregates global context information and local detailed features, enhancing the overall understanding of the target and background while preserving spatial details. This structure can identify complex spatial relationships, effectively handle multiple objects and complex backgrounds, improve the detection ability for objects of different scales, and enhance the robustness against small objects and background noise. The introduction of global information enables GLSA to maintain good recognition and segmentation capabilities even when the objects are occluded or overlapping. Compared with the traditional multi - scale feature pyramid, GLSA pays more attention to the collaborative processing of global and local information, so it has more advantages in dealing with complex backgrounds and multi - object detection tasks. Figure 4 shows the network structure diagram of GLSA.

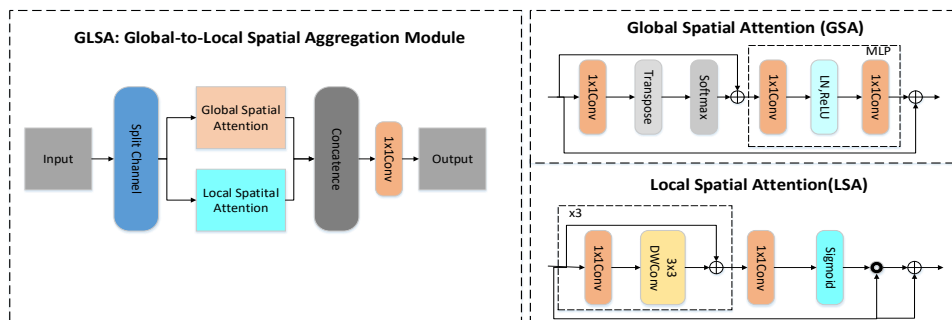


Figure 4. The Network Structure Diagram of the GLSA

## 2.5. Integration of the BiFPN (Bidirectional Feature Pyramid Network) Feature Pyramid Network

The YOLOv8 network architecture employs the FPN+PAN [15] (Feature Pyramid Network) framework. In FPN, the feature information is transmitted in a top-down manner with lateral connections [16]. The bottom-layer input of PAN is entirely reliant on the feature information that has been processed by FPN. This over-dependence may cause a partial loss of the original feature information acquired from the backbone network, thereby having an impact on the detection accuracy.

Building upon the FPN structure, BiFPN removes the nodes that have only a single input edge. It utilizes bidirectional cross-scale connections and weighted feature fusion techniques. By doing so, it reinforces the contribution of each node in the process of fusing diverse features to the overall feature network, thereby attaining a more optimal equilibrium between accuracy and efficiency, as depicted in Figure 5. With regard to the multi-object road obstacle detection task in this research, this paper incorporates the weighted bidirectional feature pyramid BiFPN[17] architecture. The YOLOv8 network that is integrated with BiFPN enables better connectivity among the various levels of the feature pyramid. This enhanced connectivity is conducive to providing more comprehensive global and more crucial feature information, and it also boosts the detection capability for obstacles of different scales.

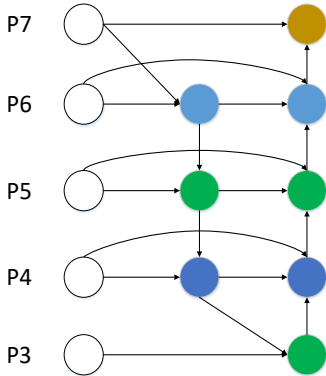


Figure 5. The Structure Diagram of the BiFPN

## 2.6. Improvement of the Loss Function

The loss function applied in YOLOv8 is CIoU [18]. Through experimental investigations, it has been discovered that the CIoU loss function exhibits the same value under varying prediction situations. In the course of road obstacle detection, this situation leads to a decrease in both the convergence speed and the accuracy of the model. Consequently, in this paper, the MPDIoU[19] loss function is employed to substitute the CIoU loss function for the purpose of optimizing the network. The calculation formulas of the MPDIoU loss function are presented as equations (1)-(3):

$$\text{MPDIoU} = 1 - \text{IoU} + \frac{d_1^2}{h^2 + w^2} + \frac{d_2^2}{h^2 + w^2} \quad (1)$$

$$d_1^2 = (x_1^{\text{pred}} - x_1^{\text{gt}})^2 + (y_1^{\text{pred}} - y_1^{\text{gt}})^2 \quad (2)$$

$$d_2^2 = (x_2^{\text{pred}} - x_2^{\text{gt}})^2 + (y_2^{\text{pred}} - y_2^{\text{gt}})^2 \quad (3)$$

## 3. Design Implementation and Result Analysis

### 3.1. Datasets and Experimental Environment

To address the scarcity of obstacle samples in existing road obstacle datasets, this paper gathers image data from the Internet, annotates it with Labelimg, and processes it. The dataset is expanded to 6,246 images, split into training, testing, and validation sets at a 7:2:1 ratio, and named MFObstacle.

The classifications of the obstacles and the quantity of images following data augmentation are presented in Table 1:

Table 1. Categories and Quantities of Obstacles

ID	Categories of Obstacles	Number
0	rockfall	950 pictures
1	person	1255 pictures
2	car	1320 pictures
3	traffic cone	1155 pictures
4	pothole	1566 pictures
	total	6246 pictures

Each experiment mentioned in this paper was performed in the host environment that is depicted in Table 2:

Table 2. Experimental Training Environment

Project	Type
GPU	NVIDIA GeForce RTX 3090 * 1
CPU	14 Xeon(R) Gold 6330
Algorithm Framework	miniconda: cuda12.2-cudnn8+Pytorch2.2.2
Operating System	ubuntu22.04

### 3.2. Comparative Experiments on the Detection Performance of Different Networks

In order to verify the superiority of the algorithm in this paper for multi-objective road obstacle detection, and at the same time verify that the self-made MFObstacle dataset in this paper has good generalization performance [20], under the same experimental environment, the improved algorithm proposed in this paper is compared with the current mainstream object detection algorithms. The comparison involves algorithmic networks such as Faster-RCNN, SSD [21], YOLOv5, and YOLOv6[22] for the detection task. The experimental outcomes are presented in Table 3.

1) When the number of parameters and the computational complexity of the YOLOv5 model are lower than those of other compared networks, its P and mAP are superior to those of the Faster R-CNN, YOLOv6, YOLOv7-tiny, and SSD networks.

2) The improved model introduced in this paper has significant and decisive superiority over YOLOv5 in several crucial dimensions, including Precision (P), Recall (R), Mean Average Precision (mAP), and Frames Per Second (FPS). In detail, the P value of the improved model is 0.9 percentage points higher than that of YOLOv5, the R value is increased by 3.7 percentage points, the mAP value rises by 1.9 percentage points, and the FPS is enhanced by 2.52 frames per second. Furthermore, this improved model outperforms two-stage detection models like Faster R-CNN and SSD

across all relevant performance indicators.

It is evident that the algorithm presented in this paper exhibits a higher level of detection accuracy and speed.

Moreover, its overall performance outshines that of the current mainstream detection algorithms.

**Table 3.** Comparison of Detection Performance of Different Networks

Base model	P%	R%	AP@0.5/%	Params/106	GFLOPs/G	FPS
Faster R-CNN	80.1	79.2	80.2	28.09	69.1	32.1
SSD	80.6	81.3	81.6	12.33	23.3	45.56
YOLOv5	83.7	80.2	83.8	2.3	7.2	94.33
YOLOv6	82.9	79.3	82.5	4.22	11.9	85.65
YOLOv7-tiny	81.4	80.5	82.3	7.24	10.8	82.42
YOLOv8n	82.3	81.8	83.5	3.06	8.1	95.37
Ours	84.6	83.9	85.7	2.13	7.6	96.85

### 3.3. Comparative Ablation Experiments

To further verify how well each of the enhanced methods proposed in this paper works, ablation experiments were performed on the relevant modules. The experimental results are shown in Table 4.

It can be known from the results of the ablation experiments in Table 4:

Upgrading C2f to C2f\_FEMA in the backbone boosts model R by 0.2% and FPS by 2.96 frames/s, reducing parameters by 34.3%. Adding GLSA improves P by 2.2% and mAP by 1%. Integrating BiFPN raises P and mAP by 0.5% and 0.2%, respectively, while replacing MPDIoU increases R

and mAP by 0.4% and 0.1%. Comparing Experiment ① and ⑥, replacing the backbone with C2f\_FEMA and adding GLSA enhances P by 1.6%, mAP by 1.2%, and FPS by 1.24 frames/s. In Experiment ⑦, integrating BiFPN increases computation but, compared to Experiment ⑥, cuts parameters by 16% while improving P, R, and mAP. Experiment ⑧, combining all four improvements, outperforms the base model with gains of 2.3% in P, 2.1% in R, 2.2% in mAP50, and 1.48 frames/s in FPS, and a 30.4% parameter reduction. Analyzing Table 4's ablation results validates the effectiveness of each proposed module.

**Table 4.** Results of the Ablation Experiments

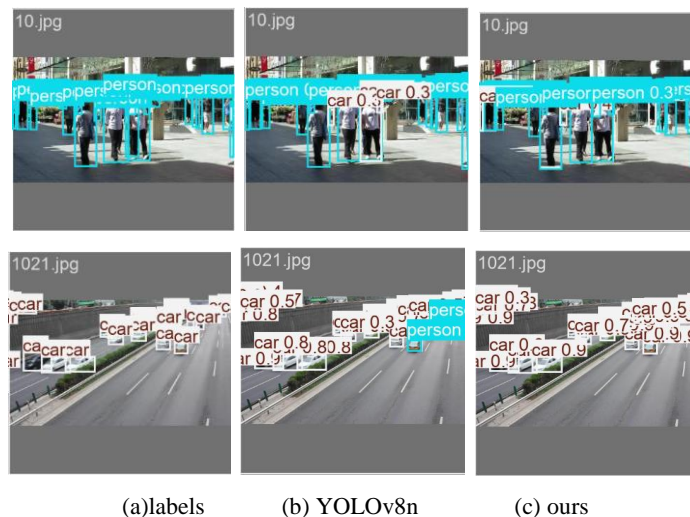
ID	C2f_FEMA	GLSA	BiFPN	MPDIoU	P%	R%	AP@0.5/%	Params/106	GFLOPs/G	FPS
1	/	/	/	/	82.3	81.8	83.5	3.06	8.1	95.37
2	√	/	/	/	82.5	82.3	83.1	2.01	7.2	98.33
3	/	√	/	/	84.5	82.6	84.5	3.22	7.9	95.31
4	/	/	√	/	82.8	82.6	83.7	2.78	7.5	92.46
5	/	/	/	√	82.4	82.2	83.6	3.06	8.1	96.53
6	√	√	/	/	83.9	83.5	84.7	2.55	7.4	97.61
7	√	√	√	/	84.1	83.6	85.4	2.13	7.6	97.01
8	√	√	√	√	84.6	83.9	85.7	2.13	7.6	96.85

### 3.4. Algorithm Verification

In order to vividly display the detection performance of the proposed network, this research makes a comparison between the obstacle detection results of YOLOv8n and the improved network, which is demonstrated in Figure 6.

The first - row image comparison reveals that the improved

algorithm significantly boosts obstacle detection classification accuracy and excels in multi - object recognition. The second - row results show that the improved algorithm can detect distant obstacles missed by YOLOv8n, demonstrating its ability to address misidentification of small, far - away obstacles and enhance confidence levels.



**Figure 6.** Comparison of Detection Results before and after Improvement

## 4. Conclusion

In this paper, an enhanced YOLOv8 obstacle object detection method is put forward. The C2f\_FEMA improved backbone network has been developed with the aim of accelerating the model's detection speed. By incorporating the GLSA spatial fusion module and integrating BiFPN, the network's performance in extracting obstacle features is elevated. Besides, MPDIoU is adopted as the loss function to fine-tune the network's bounding box regression capability. The experimental results of the improved model on the self-created MFObstacle dataset indicate that, in contrast to the initial YOLOv8, the method presented in this paper leads to an increase in the accuracy, mean average precision, recall rate, and FPS indices by 2.3 percentage points, 2.1 percentage points, 2.2 percentage points, and 1.48 frames per second respectively. Simultaneously, it also reduces the quantity of model parameters. The proposed method can well handle the obstacle object detection task in the multi-objective road scene. In the future, considering that occluded obstacle objects are likely to appear in complex road scenes, resulting in pixel overlap problems, further improvements will be made to address such issues in subsequent research

## References

- [1] Zhang X Y, Gao H B, Zhao J F, et al. A review of deep learning based autonomous driving technologies[J]. *Journal of Tsinghua University (Natural Science Edition)*, 2018, 58(004): 438-444.
- [2] Ran Z S, Li R and He S. Road obstacle detection algorithm based on improved YOLOv5s[J]. *Electric Measurement Technology*, 2024, 46(22): 177-185.
- [3] GIRSHICK R. Fast R-CNN[C]. *Proceedings of the IEEE International conference on computer vision*, Santiago, Chile, 2015:1440-1448.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Montréal, Canada, 2015:91-99.
- [5] HUU P N, THI Q P and QUYNH P T. Proposing lane and obstacle detection algorithm using YOLO to control self-driving cars on advanced networks [J]. *Advances in Multimedia*, 2022: 3425295.
- [6] CAO J, SONG C, PENG S, et al. Pedestrian detection algorithm for intelligent vehicles in complex scenarios[J]. *Sensors (Switzerland)*, 2020, 20(13): 3646.
- [7] SHIVAYOGI A B, DHARMENDRA N C M, RAMAKRISHNA A M, et al. Real-time traffic sign recognition using deep learning [J]. *Pertanika Journal of Science and Technology*, 2023, 31(1): 137-148.
- [8] WANG Y R, YANG G P, GUO J B. Vehicle detection in surveillance videos based on YOLOv5 lightweight network [J]. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 2022, 70(6): 143644.
- [9] HUANG P, WANG S, CHEN J, et al. Lightweight model for pavement defect detection based on improved YOLOv7 [J]. *Sensors*, 2023, 23(16): 7112.
- [10] S. A. A. Qadri, N. -F. Huang, T. M. Wani and S. A. Bhat, "Plant Disease Detection and Segmentation using End-to-End YOLOv8: A Comprehensive Approach," 2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2023, pp. 155-160, doi:10.1109/ICCSCE58721.2023.10237169.
- [11] T. Hu, D. Zhuang, J. Qiu and L. Zheng, "Improved YOLOv8 Algorithm with C2f-DCNv3 and Shuffle Attention for Detection of Coal Shearer Drum Teeth," 2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE), Guangzhou, China, 2024, pp. 1019-1022, doi: 10.1109/NNICE61279.2024.10498575.
- [12] Z. Li, C. Jiang and Z. Li, "An Insulator Location and Defect Detection Method Based on Improved YOLOv8," in *IEEE Access*, vol. 12, pp. 106781-106792, 2024, doi: 10.1109/ACCESS.2024.3436919.
- [13] L. Nannan and W. Liuai, "Research on Helmet Wearing Detection Based on Improved YOLOv8 Algorithm," 2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2023, pp. 910-917, doi: 10.1109/ITAIC58329.2023.10408800.
- [14] X. Hu, P. Zhang, Q. Zhang and F. Yuan, "GLSANet: Global-Local Self-Attention Network for Remote Sensing Image Semantic Segmentation," in *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023, Art no. 6000105, doi: 10.1109/LGRS.2023.3235117.
- [15] Y. G. Aadarsh and G. Singh, "Comparing UNet, UNet++, FPN, PAN and Deeplabv3+ for Gastrointestinal Tract Disease Detection," 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT), Bengaluru, India, 2023, pp. 1-7, doi: 10.1109/EASCT59475.2023.10392883.
- [16] Xiaohui Liu, Yuanfu Zhao, Liyan Liu, Xiaofeng Jin, Chunfang Wang and Yue Zhao, "The analyze and design of low FPN double delta sampling circuit for CMOS image sensor," 2013 IEEE International Conference of Electron Devices and Solid-state Circuits, Hong Kong, 2013, pp. 1-2, doi: 10.1109/EDSSC.2013.6628186.
- [17] WU Y, TANG S, ZHANG S, et al. An enhanced feature pyramid object detection network for autonomous driving [J]. 2019, 9(20): 4363.
- [18] TAN M, PANG R, LE Q V, et al. Efficientdet: scalable and efficient object detection [C]. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 13-19, 2020, pp. 10781-10790.
- [19] J. Ou and Y. Shen, "Underwater Target Detection Based on Improved YOLOv7 Algorithm With BiFusion Neck Structure and MPDIoU Loss Function," in *IEEE Access*, vol. 12, pp. 105165-105177, 2024, doi: 10.1109/ACCESS.2024.3436073.
- [20] Y. Su, B. Cheng and Y. Cai, "Detection and Recognition of Traditional Chinese Medicine Slice Based on YOLOv8," 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT), Qingdao, China, 2023, pp. 214-217, doi: 10.1109/ICEICT57916.2023.10245026.