

Research on Sea Surface Ship Detection Algorithm Based on Improved YOLOv8n

Yuxin Sun

Harbin Engineering University, Heilongjiang, China

Abstract: Aiming at the problems of complex weather interference, small targets easily obscured by noise, target occlusion, and category imbalance in marine object detection for unmanned surface vehicles (USVs), this paper proposes an improved YOLOv8n-based detection algorithm for marine ships. Taking the lightweight YOLOv8n model as the baseline, the algorithm enhances detection performance in complex marine scenarios through three major strategies: reconstructing an adaptive-weight Focal Loss, embedding the CBAM hybrid-domain attention mechanism, and designing the C2f-DA feature extraction module. Experimental results on a self-built marine object dataset and the Singapore Marine Dataset (SMD) show that, compared with the original YOLOv8n, the improved algorithm increases mAP@0.5 by 2.57% and 1.5% respectively. It exhibits significantly superior detection accuracy and generalization ability for small targets, occluded targets, and complex weather conditions over the baseline model and other mainstream lightweight detection algorithms, while maintaining lightweight characteristics suitable for embedded deployment on unmanned surface vehicles.

Keywords: Object Detection; YOLOv8n; Attention Mechanism; Feature Enhancement; Class Imbalance

1. Introduction

Sea surface obstacle detection is a core technical support for USV autonomous navigation and maritime security tasks, and its performance directly determines the safety and reliability of USV operations. Single-stage object detection algorithms of the YOLO series have remarkable efficiency advantages due to the absence of proposal box generation. As a lightweight branch, YOLOv8n has become the preferred model for USV embedded platforms with its

small size, low parameter count and fast inference speed [1].

However, actual maritime detection scenarios face challenges such as complex weather (fog, rain, nighttime) leading to degraded image quality, small obstacle targets easily disturbed by marine background noise, target occlusion, scale variations and dataset class imbalance, which severely impair YOLOv8n's detection accuracy. Small target detection difficulty is one of the most prominent problems in sea surface object detection. In maritime scenarios, some obstacle targets are small in size, especially in large-scale Synthetic Aperture Radar (SAR) or satellite images. Small obstacle targets have indistinct feature information and are easily confused with the sea surface background, resulting in a high miss detection rate. To address this issue, scholars have proposed various optimization algorithms. Zhu M et al. proposed an anchor-free detection method based on the fully convolutional one-stage object detector (FCOS), which eliminates the limitation of anchor boxes on small target detection by performing pixel-wise obstacle detection, effectively avoiding miss detection of small obstacles. Combined with the statistical characteristics of SAR obstacles, the method redesigns sample definitions to further improve adaptability for small target detection [2]. Gong Y et al. proposed an improved SSPNet, evolved from a Feature Pyramid Network (FPN), integrating a Context Attention Module (CAM), a Scale Enhancement Module (SEM) and a Scale Selection Module (SSM). The CAM introduces attention heat maps to enhance focus on small target features; the SEM uses residual modules to enable the network to focus on targets of specific scales; the SSM integrates deep semantic features into shallow features to compensate for insufficient small target features. Meanwhile, a weighted negative sampling strategy is proposed to select more representative samples, significantly improving detection performance for small obstacle targets

[3][4].

Obstacle occlusion is another key factor affecting sea surface object detection accuracy. In maritime scenarios, obstacles often occlude each other, leading to incomplete target features, which prevents detection models from accurately identifying targets and causes miss or false detection. To address occlusion, Zhao Z et al. proposed the enhanced convolutional neural network VarifocalNet*, which introduces a deformable convolution module to enhance the model's feature extraction capability for occluded targets, effectively capturing incomplete features in occluded regions. It also redesigns the loss function, introduces a non-maximum suppression algorithm, and adopts a multi-scale prediction method to further improve recognition and localization accuracy for occluded targets, optimizing detection performance in harsh marine environments and occluded scenarios [5][6].

Harsh maritime environment interference and significant appearance differences among obstacle classes further increase the difficulty of sea surface object detection. Frequent foggy and rainy weather at sea reduces visibility, blurs obstacle features and impairs vision-based detection performance. Additionally, obstacles in real scenarios are diverse, with significant shape and size differences across classes, increasing the complexity of target recognition. Scholars have proposed targeted optimization solutions for these issues. Xu Q et al. proposed a multi-linear rescaling scheme to quantize original satellite images into 8-bit images, and introduced dilated convolution into the YOLO network to effectively extract obstacle and background features. An adaptive weighting scheme is also designed to balance losses between easily detected obstacles and hard-to-detect obstacles (e.g., blurred obstacles in harsh weather, obstacles with special appearances), improving detection model robustness [7]. Shin H C et al. proposed an automatic data augmentation method, which generates obstacle position information and training data automatically by extracting foreground obstacle masks and combining them with new backgrounds. This expands scene coverage of the dataset, alleviates training bias caused by insufficient harsh environment samples and obstacle class differences, and enhances the model's adaptability to different scenarios and obstacle classes [8].

In summary, current sea surface object detection faces core challenges including small target miss detection, obstacle occlusion, harsh environment interference and large appearance differences among obstacle classes. Existing research has alleviated these problems to some extent through various optimized algorithms, but further optimization is still needed to adapt to complex and variable real maritime scenarios. This paper conducts further research and improvement on obstacle detection technology targeting these existing problems.

2. Introduction to the Baseline YOLOv8n Algorithm

As the core lightweight branch of the YOLO series, YOLOv8n adopts a fully convolutional neural network architecture following a four-stage structure: Input, Backbone Network, Feature Fusion Network and Detection Head [9].

2.1 Backbone Network

The backbone network of YOLOv8n inherits and optimizes the CSPDarknet structure of YOLOv5. Its core improvement is replacing the original C3 module with a C2f module, enhancing feature extraction capability while effectively reducing model computational complexity. The basic unit of the backbone network is CBS (Conv-BatchNorm-SiLU), which replaces the traditional LeakyReLU with the SiLU activation function to alleviate gradient vanishing and improve feature expression ability of deep networks.

The C2f module adopts a Cross Stage Partial (CSP) design, consisting of three Conv modules, several Bottleneck modules and one Concat module. Information flow is divided into a main branch and a feature branch: after adjusting channels via the first Conv module, part of the input feature map directly enters the main branch as a residual edge, while the other part enters the Bottleneck series branch for deep feature mining. Features from the two branches are concatenated via Concat, then integrated and optimized through a Conv module to output the final feature map. This module enables effective gradient propagation, balancing feature extraction capability and operational efficiency. The backbone network ultimately outputs three groups of feature maps at different scales, corresponding to small, medium and large target features respectively, providing basic feature support for subsequent feature fusion and object

detection. The scale and channel number of output features are rigorously designed to meet lightweight deployment requirements.

2.2 Feature Fusion Network

The feature fusion network of YOLOv8n adopts the FPN+PAN feature pyramid structure inherited from YOLOv5. Its core function is to integrate multi-scale features output by the backbone network, realizing complementary fusion of shallow detailed features and deep semantic features, and improving the model's detection capability for obstacles of different scales.

The FPN adopts a top-down path to transfer high-level semantic features downward and fuse them with shallow detailed features, improving small target detection accuracy. Based on the FPN, the PAN adds a bottom-up path aggregation to strengthen interaction between deep and shallow features, further enhancing multi-scale feature fusion performance. Compared with YOLOv5, YOLOv8n uniformly replaces C3 modules in the Neck layer with C2f modules, consistent with the improvement idea of the backbone network, ensuring consistency and efficiency in feature fusion.

2.3 Detection Head

The detection head is the core of object detection algorithms, responsible for target localization and class recognition on fused feature maps. YOLOv8n adopts an Anchor-Free architecture and a Decoupled Head, completely abandoning the Anchor-Based architecture and coupled detection head design of previous generations, which is the key to its performance improvement.

2.4 Loss Function

The loss function of YOLOv8n consists of classification loss and regression loss, guiding parameter optimization during model training:

1) **Classification Loss:** Binary Cross-Entropy with Logits Loss (BCE With Logits Loss) is adopted [10], which addresses class imbalance and avoids computational redundancy caused by the SoftMax function. Its calculation formula is:

$$Loss_{BCE} = -y \log p - (1-y) \log (1-p) \quad (1)$$

Where: y denotes the label corresponding to the input sample, taking a value of 1 for positive samples and 0 for negative samples; p denotes the predicted probability of the input sample being a positive sample.

2) **Regression Loss:** A combination of Distribution Focal Loss (DFL) and CIoU Loss is used. DFL models bounding box coordinates as a probability distribution to improve localization accuracy for small and overlapping targets; CIoU Loss builds on IoU Loss by introducing the center distance and aspect ratio of bounding boxes, enhancing the stability and accuracy of bounding box regression [11]. Its core calculation formula is:

$$I_{CIoU} = 1 - IoU + \frac{d^2}{c^2} + \alpha v \quad (2)$$

Where: IoU denotes the Intersection over Union between the predicted box and the ground truth box; d denotes the Euclidean distance between the centers of the predicted box and the ground truth box; c denotes the side length of the minimum enclosing rectangle of the predicted box and the ground truth box; α is a weight parameter, and v denotes the aspect ratio similarity of the two boxes.

3) **Label Assignment:** A Task-aligned strategy is adopted, selecting anchor points near the center of the ground truth box as positive samples, and filtering low-confidence anchor points via Non-Maximum Suppression (NMS). Anchor point alignment is determined by both classification score and Intersection over Union.

3. Design of the Improved YOLOv8n Detection Algorithm

The improved algorithm retains the four-stage structure of YOLOv8n (Input, Backbone Network, Feature Fusion Network, Detection Head), with core improvements focusing on backbone network feature enhancement and loss function optimization, ensuring compatibility between improved modules and the original network channels while maintaining lightweight characteristics.

3.1 Design of Adaptive Weight Focal Loss

The original YOLOv8n uses BCE With Logits Loss as the classification loss, which is prone to training domination by dominant classes in maritime datasets with class imbalance, diluting gradient updates for small-sample targets (e.g., buoys, seabirds). Meanwhile, traditional Focal Loss cannot directly adapt to the three-dimensional classification output tensor of YOLOv8n, easily leading to gradient propagation failure.

This paper designs an adaptive weight Focal Loss, addressing class imbalance and

small-sample learning issues through dual mechanisms of focusing coefficient and class balance coefficient. Its mathematical expression is:

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) \quad (3)$$

Where: p_t denotes the model's predicted probability of the correct class; γ is a focusing parameter that suppresses loss from easy samples and strengthens learning for hard samples; α_t is a class balance coefficient that alleviates positive-negative sample imbalance.

To adapt to the YOLOv8n framework, an inheritance rewriting strategy is adopted: a custom Weight Focal Loss class inherits the native loss base class, rewriting only the classification loss calculation logic while retaining the original regression loss and confidence loss, ensuring compatibility with model optimizers, learning rate schedulers and other modules. Prediction probabilities are derived via exponential transformation, and class weights are introduced for loss calculation without altering tensor shapes or gradient propagation paths.

3.2 Embedding of Mixed-Domain Attention Mechanism

The convolutional operations of YOLOv8n's backbone network have limited receptive fields and lack adaptive focusing capability for key channel and spatial information [12], easily causing feature confusion in complex sea surface backgrounds. This paper introduces the CBAM attention module, embedding it after specific convolutional layers of the backbone network. It balances channel and spatial attention, enhances target features and suppresses background noise, with a small increase in parameter count that does not affect the model's lightweight characteristics.

CBAM operates serially via a Channel Attention Module (CAM) and a Spatial Attention Module (SAM). The module first performs average pooling and max pooling on the input feature map, then learns channel relationships via a Multi-Layer Perceptron (MLP) to generate channel attention weights. It subsequently pools the channel-weighted feature map along the channel dimension, extracts spatial information via a 7×7 convolution to generate spatial attention weights, and finally completes dual weighted optimization of features. The calculation formulas are [13]:

$$F' = M_C \otimes F \quad (4)$$

$$F'' = M_S \otimes F' \quad (5)$$

Where: M_C denotes the channel attention map, M_S denotes the spatial attention map, and \otimes denotes element-wise multiplication.

3.3 Design of C2f-DA Feature Extraction Module

To enhance the model's multi-scale target feature extraction capability, a C2f-DA module integrating dynamic large-kernel attention is proposed to replace the native C2f module. This module retains the split-fusion structure of C2f to maintain gradient propagation advantages, only improving the internal Bottleneck structure: fixed receptive field convolutions are replaced with deformable attention units, which autonomously learn spatial offsets to focus on key regions such as target edges and contours, enhancing feature representation capability for irregular and small targets.

Forward propagation process of the C2f-DA module: the input feature is unified in channels via a 1×1 convolution, then split into a residual branch and a feature extraction branch. The feature extraction branch performs deep feature mining via multi-layer Bottleneck-DA, and the final feature is concatenated with the residual branch feature, integrated and output via a 1×1 convolution [14]. This design is highly compatible with the native C2f, enabling direct replacement with only a slight increase in computational cost, achieving a balance between accuracy and speed.

3.4 Overall Architecture of the Improved Algorithm

The overall structure of the improved algorithm is shown in Figure 1. The input layer optimizes data augmentation strategies (Mosaic [15], HSV enhancement [16], geometric transformation enhancement, flip enhancement, MixUp [17], etc.) to alleviate small-sample and single-scene issues; the backbone network embeds the CBAM module and replaces C2f with C2f-DA to strengthen feature extraction; the loss function adopts adaptive weight Focal Loss to optimize classification accuracy; the feature fusion and detection head retain the native structure to ensure model lightweight and inference efficiency.

4. Experimental Configuration and Result

Analysis

4.1 Experimental Setup

4.1.1 Experimental Environment

All experiments are conducted in a GPU environment. To ensure rigor and scientificity of data, the control variable method is strictly followed throughout the experiments with unified hardware and software configurations, as detailed in Table 1.

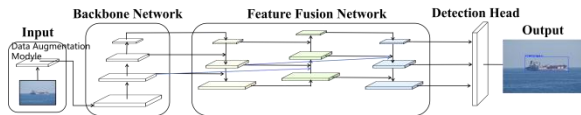


Figure 1. Overall Structure of the Improved YOLOv8n Algorithm

Table 1. Experimental Configuration Parameters

Type	Configuration
CPU	Intel Xeon Gold 6240R (Dual)
GPU	NVIDIA GeForce RTX 3090
Software Environment	CUDA 12.8, Python 3.8.19, Torch 2.0.1, ultralytics 8.4.6
Training Parameters	Batch size: 16, Epochs: 150, Early stopping patience: 15, Initial learning rate: 0.01, Optimizer: SGD

4.1.2 Dataset Introduction

To ensure comprehensiveness and scientificity of experimental results, the model’s detection accuracy is validated on the public Singapore Maritime Dataset (SMD):

Singapore Maritime Dataset (SMD): A public benchmark dataset containing 6,761 high-resolution images covering diverse complex maritime scenarios with 10 target classes, characterized by abundant small targets and large scale variations. The dataset is split into a training set (5,408 images), validation set (676 images) and test set (677 images) at an 8:1:1 ratio to validate the algorithm’s generalization ability.

4.1.3 Evaluation Metrics

Standard object detection metrics are adopted:

Table 2. Performance Comparison Between Original and Improved Models on the Dataset

Model	mAP@0.5	mAP@0.5:0.95	Precision	Recall
YOLOv8n	0.8979	0.5825	0.8962	0.8413
YOLOv8n Plus	0.9142	0.6256	0.9134	0.8711

Further analysis shows that the improved model increases mAP@0.5 from 0.8997 to 0.9132 (a 1.5% improvement), mAP@0.5:0.95 by 7.1% to 0.6265, indicating significant improvement in small target localization accuracy. Precision and Recall increase by 2% and 3.4% respectively,

Precision, Recall, mAP@0.5 (mean average precision at IoU=0.5), and mAP@0.5:0.95 (mean average precision with IoU ranging from 0.5 to 0.95 at 0.05 intervals) to evaluate model detection performance.

4.2 Analysis of Experimental Results

4.2.1 Performance Comparison Between Original and Improved Models

The performance comparison and convergence curves of the original YOLOv8n and the improved YOLOv8n_Plus are shown in Figure 2. The curves indicate that the improved model outperforms the baseline model in both convergence speed and final accuracy, with mAP@0.5 stabilizing above 0.75 and mAP@0.5:0.95 above 0.45. Training and validation loss curves show that the improved model effectively mitigates overfitting and has stronger generalization ability.

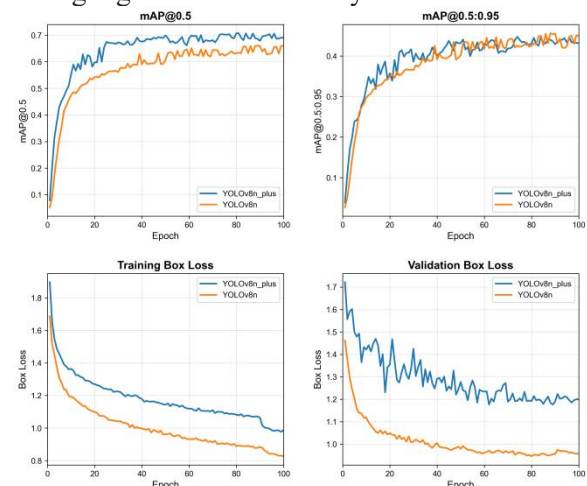


Figure 2. Accuracy and Loss Convergence Curves of Original and Improved Models

Quantitative comparison results are presented in Table 2. The improved YOLOv8n_Plus outperforms the original model in all metrics, with mAP@0.5 increased by 2.57%, and Precision and Recall reaching 0.8373 and 0.6549 respectively, achieving a balance between high precision and high recall.

with substantially enhanced generalization ability.

4.2.2 Comparison with Mainstream Algorithms

The proposed algorithm is compared with mainstream lightweight algorithms including YOLOv10n [19], YOLOv11n [20] and SSD [21],

with results shown in Table 3. YOLOv8n_Plus ranks first in mAP@0.5, Precision and Recall, with mAP@0.5 0.28% higher than YOLOv11n, verifying the effectiveness of the improved strategies. Although SSD performs well in

mAP@0.5:0.95, its Precision and Recall are inferior to the improved model, demonstrating that the proposed algorithm is more adaptable to complex maritime scenarios.

Table 3. Performance Comparison of Different Algorithms on the Dataset

Model	mAP@0.5	mAP@0.5:0.95	Precision	Recall
YOLOv8n	0.6960	0.4571	0.7858	0.5690
YOLOv10n	0.7018	0.5242	0.7749	0.6529
YOLOv11n	0.7189	0.5239	0.8038	0.5382
SSD	0.7052	0.6346	0.7869	0.6477
YOLOv8n_Plus	0.7217	0.5017	0.8373	0.6549



(a)Scenario 1



(b)Scenario 2



(c)Scenario 3



(d)Scenario 4

(1)Original image

(2)Yolov8n

(3)Yolov8n_Plus

Figure 4. Detection Effect Comparison of Original and Improved Models on the SMD Dataset

Figure 4 compares the detection effects of the original and improved models on the SMD dataset. In Scenario 1, the original YOLOv8n suffers from misdetection in the far-right region under heavy fog and low visibility, incorrectly identifying one target as two overlapping obstacles. In Scenario 2 with seemingly good visibility, YOLOv8n is easily disturbed by the background, misclassifying sky clouds as seabirds. In Scenario 3 with multiple targets, YOLOv8n fails to effectively handle overlaps, incorrectly segmenting the leftmost and rightmost cargo ships into multiple targets. In Scenario 4 with close-range targets, YOLOv8n produces two false detections in the background. In contrast, the proposed improved algorithm achieves accurate detection with no miss or false detection in all four complex scenarios,

significantly enhancing detection robustness for long-distance small sea surface targets, background noise interference and overlapping targets.

5. Conclusion and Future Work

To address complex scenario challenges in USV maritime obstacle detection, this paper proposes an improved YOLOv8n-based detection algorithm. Through collaborative optimization of adaptive weight Focal Loss, CBAM attention mechanism and C2f-DA feature extraction module, the algorithm effectively solves class imbalance, low small target detection accuracy and background noise interference issues. Experimental results show that the improved algorithm performs excellently on both the custom dataset and the SMD dataset, with

significantly enhanced detection accuracy and generalization ability compared with the original YOLOv8n and other mainstream algorithms, while maintaining lightweight characteristics suitable for USV embedded deployment.

The proposed algorithm still has limitations: a small number of false detections persist under extremely complex sea conditions (heavy fog, heavy rain), and detection accuracy for ultra-small targets can be further improved. Future work will focus on optimizing the embedding position of the attention mechanism and integrating multi-modal information (e.g., infrared images) to enhance model robustness in extreme scenarios. Additionally, model quantization and pruning techniques will be adopted to further improve inference speed, better meeting real-time detection requirements for USVs.

References

- [1] Fan Y B. Underwater Biological Target Detection Based on Deep Learning and Edge Computing [D]. Harbin: Harbin Engineering University, 2024.
- [2] Zhu M, Hu G, Zhou H, et al. A ship detection method via redesigned FCOS in large scale SAR images[J]. *Remote Sensing*, 2022, 14(5): 1153.
- [3] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 37(9): 1904-1916.
- [4] Gong Y, Zhang Z, Wen J, et al. Small ship detection of SAR images based on optimized feature pyramid and sample augmentation[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.
- [5] Zhang H, Wang Y, Dayoub F, et al. Varifocalnet: An iou-aware dense object detector[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021: 8514-8523.
- [6] Shao Z, Lyu H, Yin Y, et al. Multi-scale object detection model for autonomous ship navigation in maritime environment[J]. *Journal of Marine Science and Engineering*, 2022, 10(11): 1783.
- [7] Xu Q, Li Y, Shi Z. LMO-YOLO: A ship detection model for low-resolution optical satellite imagery[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 4117-4131.
- [8] Shin H C, Lee K I, Lee C E. Data augmentation method of object detection for deep learning in maritime image[C]. 2020 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2020: 463-466.
- [9] X. Sun, W. Cui, H. Liu, et al. CCCS-YOLO: A Method for Steel Surface Defect Detection[J]. *IAENG International Journal of Computer Science*, 2026.
- [10] J. Chen, S. Zeng, Q. Gao, et al. Underwater Organism Detection Model Integrating Dual-Path Attention Mechanism Based on YOLO-v8n[C]. 2025 IEEE 8th International Conference on Electronic Information and Communication Technology (ICEICT). Weihai, China, 2025: 226-231.
- [11] Zhang X Y, Wang M. Research on Improved YOLOv8n Lightweight Scenic Pedestrian Detection Method [J]. *Computer Engineering and Applications*, 2025, 61(2): 84-96.
- [12] J. Niu, et al. LD-YOLO: A Lightweight Dynamic Convolution-Based YOLOv8n Framework for Robust Ship Detection in SAR Imagery[J]. *IEEE Geoscience and Remote Sensing Letters*, 2026, 23: 1-5.
- [13] S. Woo, J. Park, J. Y. Lee, et al. CBAM: Convolutional Block Attention Module[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, 2018: 3-19.
- [14] Zhong K, Gong Z X, Mei X L, et al. An Improved YOLOv8-Based Face Target Detection Method [P]. *Nanjing University of Information Science & Technology*, 2025-11-11. CN202510294889.7.
- [15] Y. Li, R. Cheng, C. Zhang, et al. Dynamic mosaic algorithm for data augmentation[J]. *Mathematical Biosciences and Engineering*, 2023, 20(4): 7193-7216.
- [16] J. Zhang, H. Su, T. Zhang, et al. Multi-Scale Fusion Underwater Image Enhancement Based on HSV Color Space Equalization[J]. *Sensors (Basel)*, 2025, 25(9): 2850.
- [17] H. Alwasiti, M. Z. Yusoff. Motor Imagery Classification for Brain Computer Interface Using Deep Convolutional Neural Networks and Mixup Augmentation[J]. *IEEE Open Journal of Engineering in Medicine and Biology*, 2022, 3: 171-177.

- [18] L. Li, Z. Hou, M. Ma, et al. Spotlight on Small-Scale Ship Detection: Empowering YOLO with Advanced Techniques and a Novel Dataset[C]. M. Cho, I. Laptev, D. Tran, A. Yao, H. Zha (Eds.). Computer Vision – ACCV 2024. Lecture Notes in Computer Science, vol. 15477. Singapore, 2025: 1-18.
- [19] Ma Z, Luo Y H, Sun H, et al. SAR Image Small Ship Detection Algorithm Based on Improved YOLOv10s [J]. Modern Electronic Technology, 2026, 49(3): 59-66.
- [20] Liu X Z. Research on Small Target Detection Method Based on YOLOv11n [D]. Huainan: Anhui University of Science and Technology, 2025.
- [21] Y. Wu, C. Chen, B. Wang. Pedestrian Detection Based on Improved SSD Object Detection Algorithm[C]. International Conference on Networking and Network Applications (NaNA). Urumqi, China, 2022: 550-555.