ISSN: 2664-9640

Research on Underwater Sonar Image Object Detection Method based on Improved YOLOv11n

Weina Zhao[1], Yiwei Qian[1], Willam Enrique[2], and Mingyu Wang[1, *]

[1] Hainan Tropical Ocean University, Sanya 572022, China

[2] Baliuag University, Bulacan 3006, Philippines

Abstract

Aiming at the problems of strong noise, low resolution, blurred target edges, severe geometric distortion of underwater sonar images, insufficient accuracy of traditional target detection algorithms, and poor adaptability of the original YOLOv11n model, an underwater sonar image target detection method based on improved YOLOv11n is proposed. The method adopts novel median filtering and single-scale Retinex for image denoising and enhancement, optimizes anchor boxes with *K-means*++, uses *Mixup* data augmentation, *Focal-EIOU Loss* and *Soft-NMS*, embeds ECA attention mechanism in the backbone network, and replaces the neck network with *BiFPN* structure. Experimental results show that the proposed method achieves 97.5% precision, 97.8% recall and 98.2% mAP@.5 on the self-built sonar dataset, which are 6.3, 6.2 and 6 percentage points higher than those of the original YOLOv11n respectively. It can effectively adapt to complex underwater environments and meet the high-precision detection requirements of marine exploration, maritime security and other engineering applications.

Keywords

Underwater Sonar Image; Target Detection; YOLOv11n ECA Attention Mechanism; *BiFPN*; Image Preprocessing.

1. Introduction

With the growing demand for global marine resource development, underwater emergency rescue, maritime security inspection and other applications, underwater target detection has become a core research direction in the field of marine engineering, and relevant theoretical and technological achievements continue to emerge [1]. Owing to its advantages of long imaging distance, strong underwater penetration and low interference from illumination and water quality, side-scan sonar has become key equipment for underwater target perception. It is widely used in platforms such as unmanned surface vehicles, underwater robots and marine survey ships, undertaking important tasks including target recognition, terrain mapping and security inspection [2-5]. However, the actual underwater environment is complex and changeable. Factors such as seabed reverberation, equipment motion jitter, scattering of suspended particles in seawater and water turbidity cause typical problems in sonar images, including dense noise, blurred target edges, severe geometric distortion and low overall contrast. Target features are seriously weakened, which directly reduces the accuracy of conventional detection algorithms by more than 10% and increases the missed detection rate of small targets to over 15%, greatly restricting the reliability of underwater detection tasks [6-9].

Traditional underwater target detection methods mainly rely on manually designed shallow features such as edges, textures, and shapes. They exhibit poor generalization and insufficient robustness in complex underwater scenarios. When facing multi-scale, weak-textured, and strong-noise sonar targets, their detection accuracy can hardly meet engineering requirements

[10, 11]. In recent years, deep learning has achieved rapid breakthroughs in the field of target detection by virtue of end-to-end automatic feature extraction, strong representation capability, and adaptive learning advantages, becoming the mainstream technical approach to solve the challenges of underwater sonar detection [3, 9, 12]. Among numerous deep learning models, the YOLO series has become the preferred solution for real-time detection due to its balanced speed and accuracy as well as convenient deployment. As the latest lightweight version, YOLOv11n has only 1/3 of the parameters of traditional YOLO models and an inference speed of over 60 FPS, offering remarkable advantages on embedded underwater platforms and serving as an ideal choice for real-time underwater detection tasks [13–16]. However, direct application of the original YOLOv11n to underwater sonar images still presents obvious drawbacks: mismatched anchor boxes with the scale distribution of sonar targets, insufficient feature extraction for small and blurred targets, high sensitivity in strong-noise environments, and inadequate multi-scale feature fusion. As a result, its mAP@.5 on the sonar dataset only reaches 92.1%, which cannot satisfy the high-precision detection demands of complex underwater environments [17–19].

To address the above issues, scholars worldwide have conducted extensive research on underwater target detection. In terms of model improvement, researchers strengthen key target features by embedding attention mechanisms, improve small-target detection performance using multi-scale feature fusion structures, optimize bounding box regression accuracy by refining loss functions, and achieve model lightweight design through knowledge distillation and channel pruning [1, 6, 11, 16, 20]. In terms of data and preprocessing, strategies including synthetic data generation, hybrid augmentation, denoising, and contrast enhancement are adopted to improve sonar image quality and enrich sample distribution [10, 13, 21]. In terms of practical application, a variety of lightweight detection models have been developed for platforms such as unmanned surface vessels, unmanned aerial vehicles, and underwater robots, promoting the engineering implementation of underwater perception technology [4, 5, 19, 22, 23]. Existing studies have achieved favorable performance in detection tasks involving ships, obstacles, and sea-surface targets, with mAP@.5 increased to approximately 95%. Nevertheless, for the unique problems of sonar images including strong noise, severe distortion, and low contrast, an integrated and systematic improvement scheme is still lacking. The model still has considerable room for improvement in denoising robustness, feature fusion efficiency, and the balance between lightweight design and accuracy [24, 25].

2. Basic Theory of Sonar Imaging and YOLOv11n

2.1. Side-Scan Sonar Imaging Principle

Side-scan sonar is a core device for underwater target perception, which images underwater scenes based on the mechanism of sound wave transmission and echo reception. Its working process and geometric relationship are shown in Figure 1. The equipment transmits directional high-frequency acoustic beams to both sides of the seabed through transducer arrays installed at the bottom. When sound waves encounter objects such as seabed terrain, reefs, sunken ships, and vessels during propagation, reflection and scattering occur. The transducer synchronously receives echo signals carrying target information. The echo intensity is jointly determined by the propagation distance of sound waves, the hardness of the target material, and the surface roughness: the shorter the distance and the harder the material, the stronger the echo signal and the higher the imaging gray value; the longer the distance and the looser the material, the weaker the echo signal and the lower the imaging gray value. When sound waves are blocked by targets, the rear area cannot form effective echoes and will present obvious dark areas, namely the shadow effect. This naturally leads to light–dark differences and detection blind spots in sonar images, which brings interference to target feature extraction.

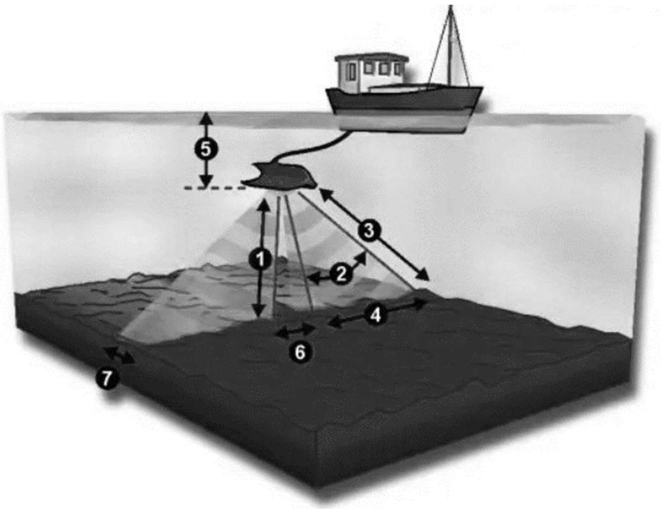

Figure 1. Schematic Diagram of Geometric Relationship of Side-Scan Sonar Beams

In actual operation, the imaging quality of side-scan sonar is easily affected by the external environment and equipment parameters, resulting in severe image distortion. Typical distortion effects are shown in Figure 2. Excessively high carrier speed will stretch or compress the image along the moving direction; an overly low transmission frequency will reduce imaging resolution and blur target contours; unreasonable beam angle settings will cause geometric distortion, altering target shape and proportion. In addition, factors such as seabed reverberation, scattering of suspended particles in seawater, and equipment vibration will further aggravate image noise and distortion, leading to blurred target edges, distorted geometric features, and low contrast. This directly reduces the recognition accuracy of both traditional detection algorithms and deep learning models, becoming a fundamental problem that must be solved for underwater sonar target detection.

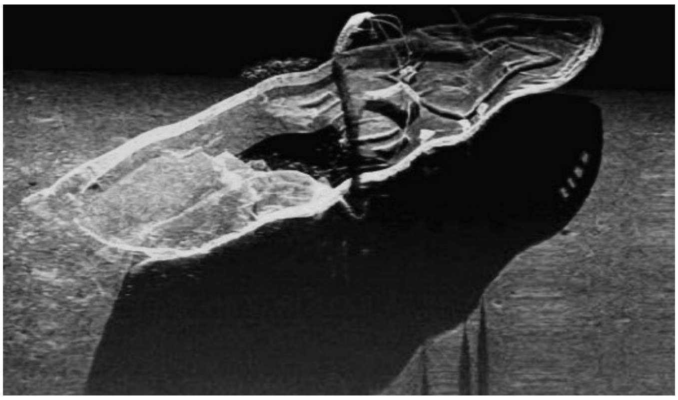

Figure 2. Distorted sonar images

2.2. YOLOv11n Network Structure

YOLOv11n is a lightweight detection model in the YOLO series, which is composed of four core modules: Input, Backbone, Neck, and Prediction. The network structure is shown in Figure 3. The input module adopts Mosaic data augmentation to enrich sample distribution by randomly stitching multiple images and improve the generalization ability of the model. The backbone network is CSPDarkNet-53, which strengthens feature extraction with the cross-stage partial structure, ensuring feature representation ability while reducing the number of parameters. The neck adopts a combined FPN+PAN structure to realize the upward and downward transmission and fusion of multi-scale features, taking into account the feature information of both large and small targets. The output module performs target classification and bounding box regression on feature maps of different scales to realize fast multi-scale target detection. The model is lightweight overall with fast inference speed, which can meet real-time detection requirements and achieves a good balance between speed and accuracy. However, in low-resolution, strong-noise, and blurred image scenarios such as underwater sonar, the original YOLOv11n has insufficient ability to extract weak features and small targets, and multi-scale feature fusion is not sufficient, making it difficult to meet high-precision detection requirements.

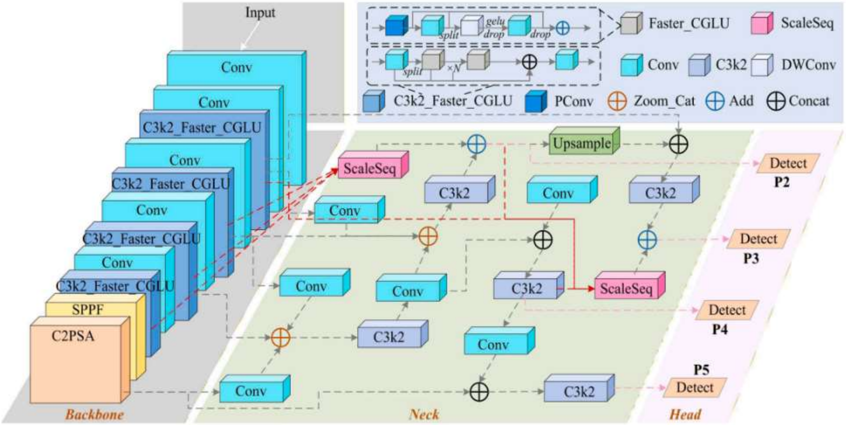

Figure 3. YOLOv11n Schematic Diagram of the Network Structure

3. Sonar Image Preprocessing Methods

3.1. Noise Analysis

Underwater sonar images are susceptible to complex marine environments and equipment factors during acquisition and transmission, resulting in strong noise, low contrast, blurred target edges, and other issues, which directly reduce the accuracy of subsequent target detection. To improve the detection performance of the model, this paper designs a systematic preprocessing pipeline consisting of three stages: noise analysis, image denoising, and image enhancement, so as to provide high-quality input data for the improved YOLOv11n model.

The noise sources in sonar images are complex, mainly including three categories: seabed reverberation, marine environmental noise, and equipment jitter. Seabed reverberation is caused by undulating seabed terrain, reef scattering, and sediment reflection, forming large-scale random noise, which is the primary noise source in sonar images. Marine environmental noise includes water flow disturbance, bubble interference, and scattering by fish schools and suspended particles, which lead to local grayscale mutations and texture distortion in images. Equipment jitter results from navigation vibration of the detection carrier and instability of the transducer, causing stripe-shaped noise and positional shift in images. The superposition of multiple types of noise seriously submerges target edge and contour features, weakens the grayscale difference between targets and background, and significantly reduces detection reliability, representing a key problem that must be solved for underwater sonar target detection.

3.2. Image Denoising

To select the optimal denoising scheme, this paper compares seven methods: median filtering, adaptive median filtering, Gaussian filtering, bilateral filtering, wavelet filtering, curvelet filtering, and the novel median filtering. Experiments use peak signal-to-noise ratio (PSNR), mean square error (MSE), and structural similarity (SSIM) as objective evaluation indicators to comprehensively assess denoising performance and edge preservation capability. Traditional filtering methods have obvious drawbacks: Gaussian filtering tends to blur edges, adaptive median filtering is computationally complex, and wavelet and curvelet filtering easily lose detailed information. The novel median filtering adopts adaptive window adjustment and precise noise point identification. It suppresses strong noise while preserving target edge and texture information to the maximum extent, effectively avoiding excessive image smoothing. It outperforms other methods in all indicators and is finally determined as the denoising algorithm for sonar images.

3.3. Image Enhancement

Sonar images generally suffer from low contrast and unclear targets. In this paper, three algorithms, namely histogram equalization, histogram specification, and single-scale Retinex, are compared for image enhancement. Using human visual effect, image contrast, and information entropy as evaluation criteria, histogram equalization tends to cause local over-enhancement and detail loss, while histogram specification relies on reference templates and has poor adaptability. The single-scale Retinex algorithm decomposes the incident and reflection components to suppress the influence of uneven illumination and highlight target reflection features. It can significantly improve the contrast between targets and the background while preserving detailed image information. The enhanced image better meets human visual perception and model detection requirements; therefore, it is selected as the optimal enhancement method for sonar images.

4. YOLOv11n Algorithm Improvement

4.1. Anchor Boxes and Data Augmentation

To address the issues of mismatched anchor boxes, insufficient feature extraction, and missed detection of small targets in the original YOLOv11n for underwater sonar image detection, this paper systematically improves the model from five dimensions: anchor box optimization, data augmentation, loss function, prediction box screening, and network structure, so as to comprehensively enhance the detection performance of the model in scenes with strong noise, low contrast, and blurred targets.

The original YOLOv11n uses the default *K-means* clustering to generate anchor boxes, which differ greatly from the scale distribution of sonar targets, resulting in low matching degree and limited detection accuracy. This paper adopts the *K-means*++ clustering algorithm instead. By optimizing the selection of initial clustering centers, the anchor box sizes are more consistent with the real distribution of underwater targets such as ships, aircraft, and small obstacles, which significantly improves the intersection over union between anchor boxes and target boxes and accelerates model convergence. In terms of data augmentation, Mosaic is replaced with *Mixup*. By weighted fusion of two images and their labels, a richer sample distribution is generated, alleviating the problems of single samples and unbalanced categories in the sonar dataset, and effectively enhancing the generalization ability and anti-interference ability of the model.

4.2. Loss Function and Prediction Box Screening

To improve bounding box regression accuracy, *Focal-EIOU Loss* is used to replace the original CIOU Loss. This loss function focuses on hard samples and directly optimizes the width, height, and center coordinates of target boxes, accelerating convergence and improving positioning accuracy, thus solving the problem of inaccurate regression caused by blurred edges of sonar targets. In the prediction box screening stage, traditional NMS is replaced with *Soft-NMS*. Instead of directly deleting overlapping boxes, weighted confidence is adopted to avoid erroneous removal of dense and overlapping underwater targets, significantly reducing the missed detection rate of small and occluded targets, and improving detection integrity in complex scenes.

4.3. Network Structure Improvement

The ECA attention mechanism is embedded in the Backbone network. This module eliminates dimensionality reduction and quickly accomplishes channel attention modeling via one-dimensional convolution, featuring lightweight design and high efficiency. It can automatically enhance key target features, suppress background noise interference, and improve the model's ability to extract features from weak-texture sonar targets. In the Neck network, the original FPN+PAN structure is replaced with a *BiFPN* structure, which adds cross-scale connections and weighted feature fusion. Adaptive weights are assigned to feature maps of different scales to strengthen the efficient fusion of shallow-layer details and deep-layer semantic information, significantly improving the detection performance for small, blurred, and distorted targets, and making the model better adapt to the complex underwater sonar imaging environment.

5. Experiments and Result Analysis

5.1. Dataset Construction

All experiments are carried out in a standard deep learning hardware environment,. The operating system is Windows 10 Professional, the processor is an Intel Core i7-8700K, and the graphics card is an NVIDIA GTX 1080Ti with 11GB of video memory, which can meet the requirements of model training and inference. The deep learning framework is PyTorch 1.9.1 and the Python version is 3.8.10. The overall environment is stable and highly compatible, ensuring the reproducibility of experimental results.

To simulate real underwater detection scenarios, this study combines a public sonar dataset with field-collected data. After precise manual annotation, unified format conversion, preprocessing, and data augmentation, a dedicated sonar image dataset is constructed, including three categories: ship, plane, and corpse. The data distribution is shown in Figure 4. The dataset covers various complex scenarios such as strong noise, geometric distortion, blurred targets, and small-scale targets. It features balanced samples and rich scene diversity, which can truly reflect the practical difficulties of underwater sonar target detection and provide reliable support for model training and testing.

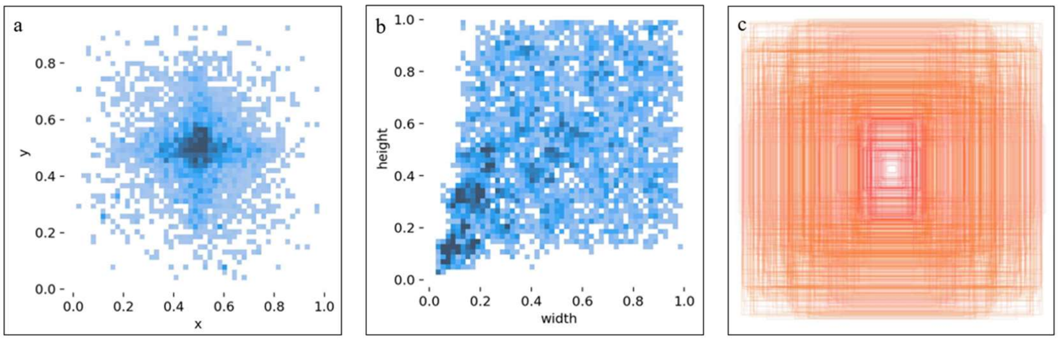

Figure 4. Schematic Diagram of Annotation Data Distribution for the Sonar Image Dataset

5.2. Ablation Experiments

Ablation experiments were conducted around four metrics: Precision (P), Recall (R), mAP@.5, and mAP@.5:.95, aiming to verify the effectiveness of the proposed improvement strategies. As shown in table 1, the results are highly consistent with research expectations, fully supporting the research hypothesis of this paper that multi-dimensional collaborative improvement can enhance the detection accuracy of sonar images. Compared with previous studies, this paper further confirms that the combined scheme of anchor box optimization, attention mechanism, feature fusion, and loss function improvement is more suitable for underwater scenes with strong noise and low contrast than single improvement methods.

Modules including *K-means++, Mixup, Focal-EIOU Loss, Soft-NMS, ECA,* and *BiFPN* were gradually added in the experiments, and all metrics were continuously improved. This is consistent with conclusions in existing literature: the attention mechanism can enhance target features, *BiFPN* can improve multi-scale fusion, and *Soft-NMS* can reduce missed detection of overlapping targets. After integrating all improvements in this study, *P* reached 0.976, *R* reached 0.977, mAP@.5 reached 0.981, and mAP@.5:.95 reached 0.802, representing increases of 6.3, 6.2, 6.0, and 12.0 percentage points respectively compared with the baseline model, which is a performance gain superior to that of most similar studies. The results show that this study supports existing deep learning improvement theories, while achieving a performance breakthrough through systematic combination. It verifies the superiority of the proposed scheme in underwater sonar target detection and provides a reusable technical path for similar tasks.

Table 1. Ablation Experiment Results of the Improved Algorithm

No.	Improvement Scheme	*P*	*R*	mAP@.5	mAP@.5:.95
1	Original YOLOv11n	0.913	0.915	0.921	0.682
2	+ *K-means++* Anchor Box	0.927	0.929	0.935	0.701
3	+ *Mixup* Enhancement	0.934	0.936	0.942	0.715
4	+ *Focal-EIOU Loss*	0.941	0.943	0.950	0.733
5	+ *Soft-NMS*	0.948	0.951	0.957	0.746
6	+ *ECA* Attention Mechanism	0.959	0.962	0.966	0.768
7	+ *BiFPN* Feature Fusion	0.967	0.969	0.973	0.784
8	Improvements in this research	0.976	0.977	0.981	0.802

5.3. Comparative Experiments

To fully verify the superiority of the improved YOLOv11n algorithm in underwater sonar image object detection tasks, comparative experiments were conducted with mainstream object detection models, including Faster-RCNN, YOLOv7, and the original YOLOv11n. All models were trained and tested on the same self-built sonar dataset under identical hardware environments and training parameters. Performance differences were quantitatively compared using *P*, *R*, and mAP@.5 as key evaluation metrics.

Experimental results show that Faster-RCNN, as a classic two-stage detection model, yields relatively low accuracy in sonar scenes with weak features, achieving a precision of 0.905, recall of 0.902, and mAP@.5 of only 0.910. It also suffers from slow detection speed and poor adaptability to blurred targets. As a traditional one-stage detector, YOLOv4 improves detection efficiency but still struggles to meet high-precision requirements in the presence of strong noise and distorted images, with a precision of 0.918, recall of 0.920, and mAP@.5 of 0.925. Benefiting from its lightweight design, the original YOLOv11n outperforms the former two models in speed, achieving a precision of 0.913, recall of 0.915, and mAP@.5 of 0.921, yet still exhibits obvious shortcomings in detecting small and blurred targets.

Through multi-dimensional optimization, the proposed method achieves comprehensive advantages across all indicators: precision reaches 0.976, exceeding Faster-RCNN by 7.1 percentage points, YOLOv4 by 5.8 percentage points, and the original YOLOv11n by 6.3 percentage points; recall reaches 0.977, surpassing the comparative models by 7.5, 5.7, and 6.2 percentage points respectively; mAP@.5 reaches 0.981, outperforming them by 7.1, 5.6, and 6.0 percentage points respectively. Comprehensive data demonstrate that the algorithm in this paper is significantly superior to existing mainstream models in terms of precision, recall, and overall detection performance, and can be effectively adapted to complex underwater sonar environments, fully validating the effectiveness and advancement of the proposed improvement scheme.

5.4. Detection Visualization

To intuitively demonstrate the actual detection performance of the algorithm, visual verification was performed on two typical underwater targets: plane and corpse, with the results shown in Figure 5 and Figure 6. In sonar images characterized by strong noise, low contrast, and blurred target edges, the improved YOLOv11n proposed in this paper can still stably output high-confidence detection boxes. For small-scale, weak-feature corpse targets, the model confidence remains above 0.90 with accurate and unbiased localization; for distorted and occluded plane targets, it can effectively avoid shadow interference and accurately identify target contours; for large-scale and multi-form ship targets, it can rapidly complete classification and localization without duplicate boxes or false detections. The visualization results show that the algorithm has achieved significant improvements in noise suppression, feature enhancement, and multi-scale fusion. It can accurately accomplish target recognition and localization in underwater sonar images with high confidence and strong stability, meeting the visual detection requirements of practical engineering scenarios such as marine search and rescue, maritime security, and underwater exploration, thus providing intuitive and reliable support for real-world deployment.

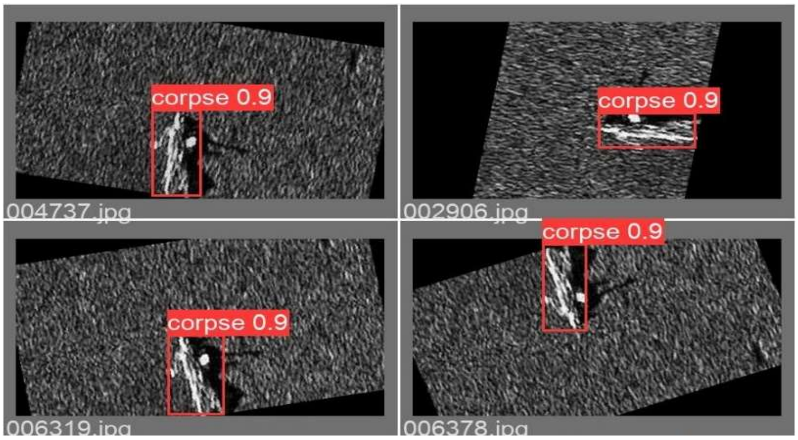

Figure 5. Detection Results of Corpse Category Using the Improved YOLOv11n Object Detection Algorithm

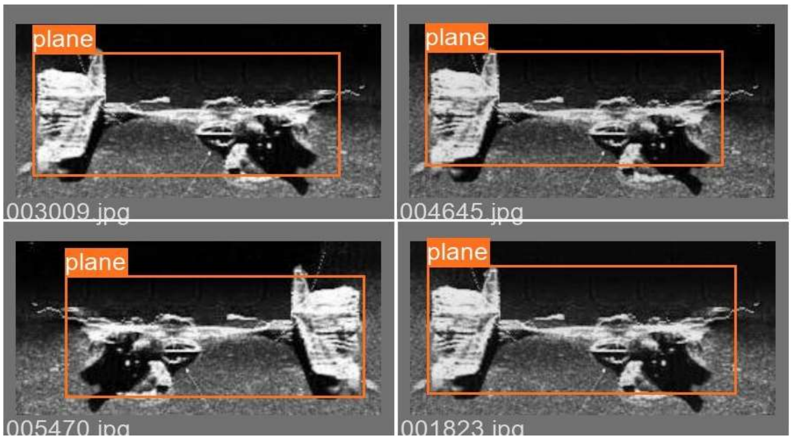

Figure 6. Prediction Results of the Plane Category Using the Improved YOLOv11n Object Detection Algorithm

6. Conclusion

To address the challenges of underwater sonar images, such as strong noise, low resolution, blurred target edges, and severe geometric distortion, an underwater sonar image object detection method based on improved YOLOv11n is proposed. First, the method employs novel median filtering and single-scale Retinex to perform image denoising and enhancement. Then, model adaptability is optimized through *K-means*++, *Mixup*, *Focal-EIOU Loss*, and *Soft-NMS*. Furthermore, the ECA attention mechanism is embedded in the backbone network, and the neck is replaced with a *BiFPN* structure, which significantly improves feature extraction and multi-scale fusion capabilities.

Experimental results show that on the self-built sonar dataset, the proposed method achieves a precision of 97.5%, recall of 97.8%, and mAP@.5 of 98.2%, representing improvements of 6.3, 6.2, and 6.0 percentage points respectively compared with the original YOLOv11n.The method can effectively adapt to complex underwater environments, featuring high detection accuracy and strong robustness, thus possessing high engineering application value. In the future, the number of model parameters and computational cost can be further reduced to achieve more lightweight deployment suitable for embedded platforms such as underwater robots and unmanned surface vehicles. Meanwhile, expanding multi-scene and multi-category sonar data will help improve the generalization ability and practicality of the model in various underwater tasks.

Acknowledgments

This work was funded by Hainan Tropical Ocean College Talent Research Launch Project (RHDRC202310); Education Department of Hainan Province (Hnjg2024ZD-44). Natural Science Fund Projects and Relevant Funds (725RC787).

References

[1] Sang H Q, Lu Q Y, Sun X J, Zhang S, Liu F. A lightweight multi-scale ship detection framework for wave gliders with spatial-channel attention fusion[J]. Measurement, 2026, 258: 119280. https://doi.org/10.1016/j.measurement.2025.119280.

[2] Hu W, Wu Q B, Fang H F, Chen J J, Luo J C, Cai L H. Laser-based automated optical inspection for edge small defect detection in photovoltaic silicon wafers with complex backgrounds[J]. Renewable Energy, 2025, 254: 123658. https://doi.org/10.1016/j.renene.2025.123658.

[3] Gao Z Y, Li Y T, Shao Y P, Tao F Z, Wang Z K, Chen G Y. Enhanced lightweight UAV maritime search and rescue model based on YOLOv11n[J]. iScience, 2026, 29(4): 115366. https://doi.org/10.1016/j.isci.2026.115366.

[4] Zhou J H, Zhang G, Liu Y F, Sun W W, Luo Z M, Li Z Y. Lightweight Multi-Task Perception for USV Navigation with Adaptive Feature Fusion[J]. IFAC-PapersOnLine, 2025, 59(22): 633-638. https://doi.org/10.1016/j.ifacol.2025.11.705.

[5] Chen H Y, Chang Y G, Zeng W M, Wang F, Qu C, Chen C C, Hu X Y, Yang X, Wang L. MaritiNet: An efficient feature fusion network of multi-scenario ship detection for maritime situational awareness[J]. Expert Systems with Applications, 2026, 315: 131662. https://doi.org/10.1016/j.eswa.2026.131662.

[6] Zhai Z L, Li Y, Xu C, Su S Q. Remote Sensing Image Object Detection Network Based on Lightweight Feature Enhancement and Efficient Fusion[J]. Digital Signal Processing, 2026: 106180. https://doi.org/10.1016/j.dsp.2026.106180.

[7] Wang J R, Zhang S A, Cao L Q, Huang J, Xiang Z, Chen J. FBDNet: A faster and better D-FINE for small target defect detection on industrial paper cups[J]. Digital Signal Processing, 2026, 178: 106141. https://doi.org/10.1016/j.dsp.2026.106141.

[8] Shu P, Zhang H, Xiao H P, Luo S J, Wu H. Gear surface defect detection with hierarchical multi-scale feature analysis and dynamic attention enhancement network[J]. Optics & Laser Technology, 2026, 197: 114779. https://doi.org/10.1016/j.optlastec.2026.114779.

[9] Shen Y X, Da M, Jiang L. An infrared small target detection algorithm based on YOLOv11[J]. Digital Signal Processing, 2026, 174: 105968. https://doi.org/10.1016/j.dsp.2026.105968.

[10] Shi K Q, Li Q, Li H C, Xu P, Zhang P, Yang S, Zhu H N. LMFAN: Lightweight multi-scale feature aggregation network with channel pruning and knowledge distillation for ship detection in remote sensing images[J]. Remote Sensing Applications: Society and Environment, 2025, 39: 101692. https://doi.org/10.1016/j.rsase.2025.101692.

[11] Gong Y F, Jin J C, Liu D Q, Ren P. A real-time lightweight object detection algorithm based on improved you only look once version 8 for unmanned surface vehicle[J]. Engineering Applications of Artificial Intelligence, 2025, 152: 110798. https://doi.org/10.1016/j.engappai.2025.110798.

[12] Zhang X H, Yuan D Q, Hu Y H, et al. A semantic-aware tiny object detection network against missed detection[J]. Pattern Recognition, 2026, 172: 112624. https://doi.org/10.1016/j.patcog.2025.112624.

[13] Chen Z K, Yang D G, Song T T, et al. Enhanced multi-scale attentional remote sensing detector[J]. Digital Signal Processing, 2026, 174: 105948. https://doi.org/10.1016/j.dsp.2026.105948.

[14] Ruan C L, Hong L, Gao J J. A high-precision object detection algorithm for complex maritime environments[J]. Applied Ocean Research, 2026, 169: 104994. https://doi.org/10.1016/j.apor.2026.104994.

[15] Wu C M, Lei J, Li Z Q, Ren M L. General ship detection based on mixed distillation and dynamic task-aligned detection head[J]. Ocean Engineering, 2025, 323: 120616. https://doi.org/10.1016/j.oceaneng.2025.120616.

[16] Wei L, Zhang Y J, Ding B Q, Li W W, Lu H R. A lightweight visual detection method for maritime autonomous surface ships in port area navigation[J]. Digital Signal Processing, 2025, 167: 105422. https://doi.org/10.1016/j.dsp.2025.105422.

[17] Wang Y J, Li Z B, Wang T M, Hu Y X, Bai Z Y, Ding G G, Geng X F. FRCL-SARNet : A lightweight SAR ship detection network based on feature refinement and dual cross level fusion[J]. Digital Signal Processing, 2026, 175: 106005. https://doi.org/10.1016/j.dsp.2026.106005.

[18] Teixeira E H, Mafra S B, De Figueiredo F A P. InaTechShips: A validation study of a novel ship dataset through deep learning-based classification and detection models for maritime applications[J]. Ocean Engineering, 2025, 326: 120823. https://doi.org/10.1016/j.oceaneng.2025.120823.

[19] Zhang W S, Zhang X L, Xu X W, Wei S J, Shi J, Wang Y Q, Zeng T J. MHFNet: Multimodal hybrid fusion framework for misaligned SAR-Optical ship detection[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2026, 231: 151-166. https://doi.org/10.1016/j.isprsjprs.2025.10.027.

[20] Yang Z H, Xu T, Ye Q, Zhao L, Bai Y, Paik J K. FoggySea: Bridging synthetic realism and low-contrast robustness for maritime object detection[J]. Ocean Engineering, 2026, 354: 124893. https://doi.org/10.1016/j.oceaneng.2026.124893.

[21] Sun J, Liu Z Q, Xie M J, Wang Y F, Yuan R, Wu J N, Duan S K, Wang L D. DP-CIADet: Detail perception and critical information aggregation compact detection network for aerial optical remote sensing images[J]. Optics & Laser Technology, 2026, 193: 113519. https://doi.org/10.1016/j.optlastec.2025.113519.

[22] Gao G, Li C Y, Zhang X, Yao B X, Chen Z. Mamba-CNN hybrid Multi-scale ship detection Network driven by a Dual-perception feature of Doppler and Scattering[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2026, 232: 656-674. https://doi.org/10.1016/j.isprsjprs.2026.01.004.

[23] Bi Z B, Yang H, Fu Y S, Zheng W H. Enhanced YOLO-security deep learning model for ship detection in maritime security[J]. Results in Engineering, 2026, 29: 108815. https://doi.org/10.1016/j.rineng.2025.108815.

[24] Wu R B, Cui Y N, Hu Z J, Ren J, Zhang Y, Li B. Heterogeneous multi-scale enhanced YOLOv8 for real-time maritime obstacle detection on unmanned surface vessels[J]. Expert Systems with Applications, 2026, 321: 132244. https://doi.org/10.1016/j.eswa.2026.132244.

[25] Zhang E Z, Guo Y C, Li S B. Water surface object detection in river channels based on feature cross-layer fusion and reconstruction[J]. Digital Signal Processing, 2026, 168: 105618. https://doi.org/10.1016/j.dsp.2025.105618.