2025 5th Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS) | 979-8-3315-2463-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ACCTCS66275.2025.00215

YOLOv8-BFDS: A Framework for Underwater Object Detection Based on Feature Enhancement and Fusion

Jiatong Li
Faculty of Information Science and Engineering
Ocean University of China
Qingdao, China
ljt4216@stu.ouc.edu.cn

***Abstract*—Object detection plays a critical role in various practical applications, but it is often hindered by challenges such as insufficient lighting, blurred imaging, and complex background interference. This paper proposes an enhanced YOLOv8 model, YOLOv8-BFDS, which integrates three key optimizations: the DCNv2 module dynamically adjusts the receptive field to better capture deformed and occluded objects; the E-SEModule combines channel and spatial attention mechanisms to enhance the detection of small and low-contrast objects; and the Concat_BiFPN module optimizes multi-scale feature fusion through bidirectional connections and adaptive weighting. Experimental results demonstrate significant improvements over the original YOLOv8, with precision increasing by 2.1% (from 0.95006 to 0.97011), recall by 19.1% (from 0.7985 to 0.98964), mAP50 by 14.2% (from 0.86825 to 0.99149), and mAP50-95 by 34.5% (from 0.62178 to 0.83604). These advancements demonstrate the model's superior robustness and accuracy in object detection tasks.**

Keywords—YOLOv8-BFDS, BiFPN, DSConv, attention mechanisms, multi-scale feature fusion

I. Introduction

Object detection plays a critical role in various research fields, including robotics, surveillance, and industrial applications. However, this task is often hindered by challenges such as insufficient lighting, blurred imaging, and complex background interference. These issues are particularly significant when dealing with small, low-contrast, and occluded objects, which often require advanced feature extraction and fusion techniques. Despite the ongoing advancements in detection algorithms, enhancing the accuracy and robustness of object detection models in such environments remains a significant challenge. Therefore, improving the performance of object detection models through innovative strategies and model optimization is of paramount importance in overcoming these obstacles.

With the rapid development of computer vision technologies, deep learning-based object detection algorithms have gradually become mainstream, giving rise to a variety of representative algorithms such as SSD, Faster R-CNN, and YOLO. These advancements have significantly propelled the field of object detection, and researchers have conducted extensive studies in this area. Jianqun Zhou proposed the addition of a spatial attention-based feature aggregation module, which preserves unique features by focusing on the instance parts of images, addressing the challenge of generating diverse data with limited real-world data using traditional generative models [1]. Yonghui Huang introduced an improved structural defect detection model based on YOLOv8, solving the challenges of multi-scale and lightweight design for structural defects [2]. Li Han proposed the CB-YOLOv5s algorithm for fish target detection, constructing a bidirectional channel to achieve cross-scale connections, thereby addressing the issue of low detection accuracy caused by mutual occlusion of targets and their similarity to the background in marine environments [3]. Fan Zhang introduced a contrastive learning mechanism into the generator residual blocks and the region proposal network of the object detection task, highlighting targets and reducing background interference, thus resolving the issue where enhanced images might degrade detection accuracy [4]. Tianbao Han proposed MFLNet, a lightweight camouflaged target detection method for unmanned surface vehicles based on a multi-task learning strategy, addressing the challenges posed by complex backgrounds, diverse shapes, and camouflage of targets, as well as the increasing diversity and complexity of detection scenarios [5].

Although the aforementioned methods have made significant contributions, they still face challenges such as insufficient lighting, blurred imaging, and complex background interference. Based on research into the YOLO series of algorithms, this paper proposes structural optimizations to the YOLOv8 model to address these challenges. The YOLO (You Only Look Once) series of algorithms has secured a prominent position in the field of object detection due to its real-time performance in single-stage detection. As the mainstream version of this series, YOLOv8 stands out for its simplicity, efficiency, strong real-time capability, and multi-scale detection. Its unique advantage lies in the optimization of feature extraction and network structure design, fully leveraging global contextual information and multi-task learning, thereby excelling in fast object detection and real-time applications.

YOLOv8 has demonstrated outstanding performance in land-based detection tasks. However, its performance in marine environments is still limited by various factors and faces numerous challenges. The first challenge is the complex lighting conditions. Unstable lighting and low contrast due to scattering and absorption directly impact YOLOv8's performance. Complex lighting conditions also make it difficult for drones and robots to maintain consistent detection results under dynamic lighting. Secondly, diverse background interference is another major issue. The environment is often filled with various background objects, increasing the difficulty of detecting target objects. Although YOLOv8's

979-8-3315-2463-0/25/$31.00 ©2025 IEEE
DOI 10.1109/ACCTCS66275.2025.00215

Authorized licensed use limited to: Ocean University of China. Downloaded on March 13,2026 at 11:00:08 UTC from IEEE Xplore. Restrictions apply.

convolutional neural network structure has some ability to separate the background, the false detection and missed detection rates remain high in complex backgrounds. In addition, small object detection and occlusion are common challenges in object detection. Particularly in long-distance captures, details of the target objects are often lost, leading to lower recognition rates. Finally, the scarcity of annotated data is another limiting factor. Annotated data is rare, and the cost of data collection and annotation is high, resulting in existing datasets failing to fully cover the variations in target object types and poses. This further limits the generalization ability of deep learning models.

To address the aforementioned issues, this paper proposes an improved model named YOLOv8-BFDS. The model optimizes the feature fusion and extraction process by incorporating the BiFPN and DSConv modules, thus enhancing the detection capability for multi-scale objects. Additionally, YOLOv8-BFDS improves the model's adaptability and detection accuracy in dynamic environments by effectively integrating spatial and channel attention mechanisms [6], as well as optimizing the dynamic receptive field. This results in improved robustness, particularly in the presence of complex lighting and background interference.

The contributions of this paper can be summarized as follows:

1) By dynamically adjusting the receptive field, DCNv2 enhances YOLOv8's adaptability to deformation, occlusion, and irregular objects, which is particularly effective in object detection, significantly improving the model's robustness.

2) YOLOv8 integrates the E-SEModule, based on the Squeeze-and-Excitation (SE) mechanism. Through channel and spatial attention mechanisms, this module strengthens the network's focus on key features, improving the detection accuracy of small and low-contrast objects in complex scenes.

3) By utilizing bidirectional feature fusion, feature concatenation, and adaptive weighting mechanisms, Concat_BiFPN optimizes YOLOv8's multi-scale feature fusion capabilities. This further enhances the model's performance in multi-scale object detection, especially in handling complex environments, significantly boosting the model's perception ability.

II. METHOD

A. YOLOv8 Architecture and Algorithmic Principles

As illustrated in Figure 1, YOLOv8-BFDS introduces key optimizations over YOLOv8, including the integration of DCNv2 to enhance the model's adaptability to deformations and occlusions, the incorporation of the E-SEModule based on the Squeeze-and-Excitation mechanism to improve feature focus, and the adoption of the Concat_BiFPN module for advanced multi-scale feature fusion. These modifications substantially enhance the detection performance of YOLOv8-BFDS in challenging underwater environments. The following sections provide a detailed overview of the YOLOv8-BFDS architecture and optimization strategies.

YOLOv8 (You Only Look Once Version 8) is a state-of-the-art single-stage object detection algorithm that continues the YOLO series' tradition of efficiency and accuracy. Compared to its predecessors, YOLOv8 introduces architectural enhancements, improved feature extraction, and optimized input processing, making it well-suited for complex object detection tasks [7][8].

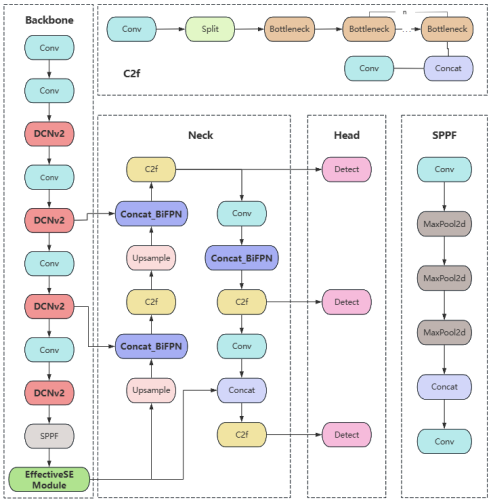


Fig. 1. The Network Architecture of the Improved YOLOv8-BFDS

The network structure consists of three main components: Backbone, Neck, and Head. The Backbone employs the C2f module, which enhances feature extraction while reducing computational cost. It also integrates the SPPF module to improve multi-scale object detection. The Neck utilizes PAFPN, a fusion of FPN and PAN, to strengthen multi-level feature representation, enhancing detection performance for small and distant objects. The Head adopts a decoupled task optimization strategy, separately refining classification and bounding box regression, along with an advanced sample assignment mechanism for better accuracy.

Additionally, YOLOv8 incorporates adaptive input processing techniques such as dynamic anchoring, Non-Maximum Suppression (NMS), and adaptive image scaling, ensuring robustness across diverse scenarios. Mosaic data augmentation further enhances small object detection by increasing data diversity [7][8][9].

With these improvements, YOLOv8 achieves high efficiency, superior detection accuracy, and adaptability across various complex environments, making it a strong foundation for further advancements in underwater object detection [10].

B. Optimization of Key Modules in YOLOv8-BFDS

1) The introduction of the DCNv2 convolution method.

To enhance YOLOv8's adaptability and robustness in underwater object detection, particularly in capturing key features under deformation and complex backgrounds, this study introduces the DCNv2 (Deformable Convolution v2) module. Unlike traditional convolution with a fixed receptive field, DCNv2 incorporates adaptive sampling offsets and weighted sampling mechanisms, dynamically adjusting the shape and size of the receptive field. This allows the convolution operation to flexibly adapt to local variations in the image, making it highly effective in extracting critical information, especially when objects undergo deformation or occlusion.

The primary advantage of DCNv2 lies in its ability to dynamically modify the receptive field based on local image features, overcoming the limitations of fixed receptive fields in traditional convolution. This significantly enhances the

Authorized licensed use limited to: Ocean University of China. Downloaded on March 13,2026 at 11:00:08 UTC from IEEE Xplore. Restrictions apply.

model's capability to capture object edges and fine details, particularly when dealing with morphologically complex or cluttered backgrounds. In underwater environments, where factors such as lighting variations and object occlusions frequently cause distortions or deformations in object appearance, the integration of DCNv2 effectively improves both robustness and detection accuracy. By incorporating the DCNv2 module into YOLOv8, the model achieves greater adaptability to underwater object detection, strengthening its ability to handle deformations and complex backgrounds, ultimately enhancing overall detection performance and precision, especially in dynamic underwater conditions.

2) The introduction of the Effective SE Module attention mechanism.

To enhance YOLOv8's detection accuracy for small and low-contrast objects in complex scenes, this study introduces the Effective Squeeze-and-Excitation Module (E-SEModule) into the YOLOv8 architecture. The E-SEModule is an optimized version of the Squeeze-and-Excitation (SE) module, which integrates both channel and spatial attention mechanisms. In the E-SEModule, global average pooling (Squeeze) is first applied to generate feature vectors for each channel. Rather than using traditional fully connected layers, convolutional layers are employed to dynamically adjust channel-wise weights, improving the representation of important features. In contrast to the conventional SE module, the E-SEModule utilizes a more efficient convolutional structure, further enhancing computational efficiency and feature representation capabilities.

Additionally, the E-SEModule incorporates a spatial weighting mechanism that applies attention across the spatial dimensions of the feature map, automatically focusing on key informative regions while suppressing irrelevant background information. This mechanism significantly improves the model's ability to highlight salient objects in complex environments, making it especially effective for detecting low-contrast and small objects. By integrating the E-SEModule into YOLOv8's feature extraction process after each convolutional block, the enhanced attention mechanism strengthens YOLOv8's sensitivity to critical features, ultimately improving detection accuracy in challenging scenarios.

3) The introduction of the Concat_BiFPN module.

To improve YOLOv8's feature fusion capability and multi-scale object detection performance in complex environments, particularly in underwater target detection tasks, this paper introduces the Concat_BiFPN module. This module combines the Bidirectional Feature Pyramid Network (BiFPN) and feature concatenation techniques to optimize the traditional feature fusion method, FPN. Through bidirectional feature fusion, feature concatenation, and adaptive weighting mechanisms, Concat_BiFPN effectively enhances the model's ability to detect multi-scale targets.

The traditional Feature Pyramid Network (FPN) performs only top-down feature fusion, which may not fully leverage the detailed information from lower-level features and is prone to losing contextual information from higher-level features. To address this issue, Concat_BiFPN introduces a bidirectional feature fusion mechanism, which supports both upsampling fusion of lower-level features with higher-level features and downsampling for information transfer from lower-level features to higher-level ones. Additionally, Concat_BiFPN employs feature concatenation instead of the traditional addition operation, a method that better preserves feature information from different levels, enhancing the network's adaptability to targets under deformation, occlusion, and complex backgrounds. Moreover, Concat_BiFPN introduces an adaptive weighting mechanism, allowing the network to dynamically adjust the fusion weights of features from each layer during training. This enables the model to focus more on the features that contribute significantly to target detection, further improving the quality and effectiveness of feature fusion.

III. Experiments

A. Environment Configuration

The experiment is based on a Windows 11 64-bit system, using PyCharm as the development tool. The Python 3.10 environment is configured with the PyTorch 2.5.0 framework, and the CUDA version is 11.8. In terms of hardware, the GPU is an NVIDIA GeForce RTX 4060, while the CPU is an Intel(R) Core(TM) i7-13650HX 2.60 GHz with 24GB of RAM.

B. Evaluation Metrics

1) Precision (P)

Precision measures the proportion of actual positive instances among all the instances predicted as positive by the model. The formula for calculation is:

$$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Positives(FP)} \quad (1)$$

Where TP refers to the number of samples that the model correctly predicts as positive, and FP refers to the number of samples that the model incorrectly predicts as positive.

2) Recall (R)

Recall measures the proportion of actual positive samples that are correctly identified by the model. The formula is as follows:

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Negatives(FN)} \quad (2)$$

Where FN (False Negative) refers to the number of samples that the model incorrectly predicts as negative.

3) mAP50

mAP50 (mean Average Precision at IoU=0.5) is the average precision at an Intersection over Union (IoU) threshold of 0.5. It reflects the model's detection accuracy under a relatively lenient matching condition. The formula for IoU is as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (3)$$

4) mAP50-95

mAP50-95 (mean Average Precision at IoU=0.5:0.05:0.95) is the mean average precision calculated over a range of Intersection over Union (IoU) thresholds from 0.5 to 0.95, with a step size of 0.05. This metric reflects the model's ability to detect objects accurately across a variety of stricter matching conditions, providing a more comprehensive evaluation of performance.

Authorized licensed use limited to: Ocean University of China. Downloaded on March 13,2026 at 11:00:08 UTC from IEEE Xplore. Restrictions apply.

C. Dataset

This study uses underwater detection video clips from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) as the dataset, which consists of 7,667 images covering 11 target categories, such as seabed organisms, plastic waste, and metal debris.

D. Comparison with Other Achievements

This study compares the performance of YOLOv5, YOLOv6, YOLOv7, YOLOv8, YOLOv9, and YOLOv11, along with YOLOv8-based models enhanced with BiFPN (Bidirectional Feature Pyramid Network) and DSConv (Depthwise Separable Convolution), on the same dataset. All models were trained using default hyperparameters for 100 epochs, with a batch size of 8. The final evaluation is based on the performance of the last epoch on the test set.

TABLE I. MODEL PERFORMANCE METRICS: YOLOV8-BFDS VS. YOLO SERIES

Model Configuration	Precision	Recall	mAP50	mAP50-95
YOLOv5	0.9592	0.86365	0.93806	0.67804
YOLOv6	0.84015	0.80368	0.81593	0.59957
YOLOv7	0.7343	0.8909	0.897	0.6745
YOLOv8	0.95006	0.7985	0.86825	0.62178
YOLOv9	0.94621	0.8083	0.85464	0.60425
YOLOv11	0.88828	0.89852	0.94707	0.69914
YOLOv8-BFDS	**0.97011**	**0.98964**	**0.99149**	**0.83604**

As illustrated in Table I, the performance metrics of each model on the same test set are summarized, covering four key evaluation indicators: Precision, Recall, mAP50, and mAP50-95.

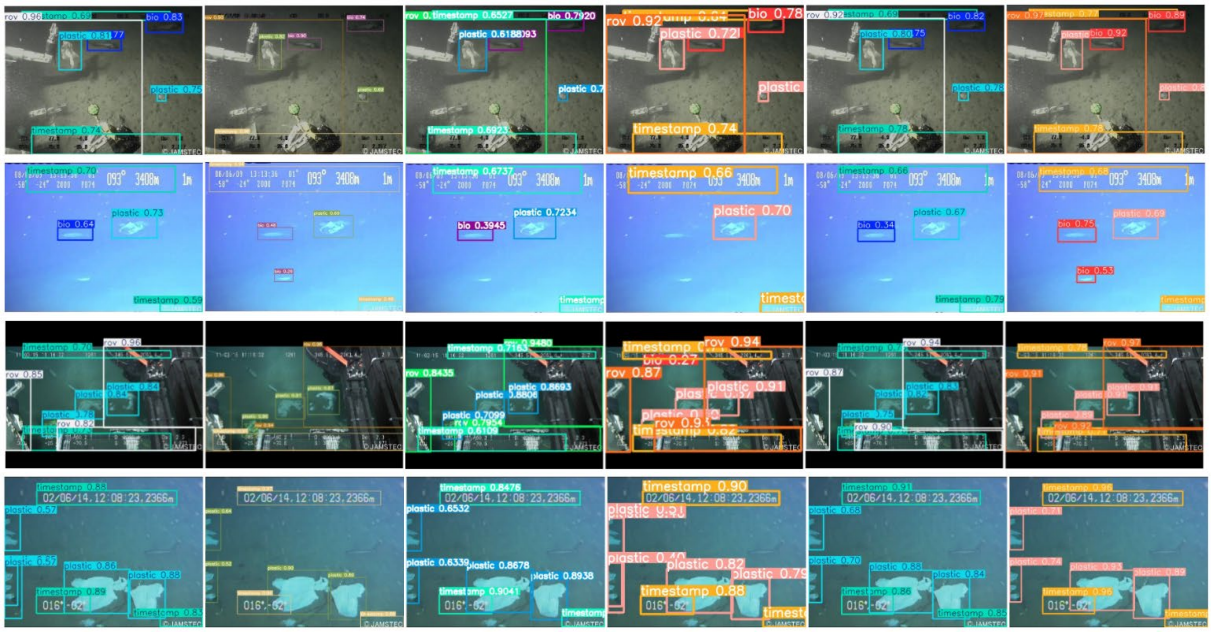
The improved model achieved a precision of 0.97011 and a recall of 0.98964, reflecting increases of 2.1% and 19.1%, respectively, compared to the original YOLOv8, which recorded a precision of 0.95006 and a recall of 0.7985. Moreover, the improved model exhibited outstanding performance in mAP50 and mAP50-95, reaching 0.99149 and 0.83604, respectively. These values represent enhancements of 14.2% and 34.5% over the original YOLOv8, which achieved an mAP50 of 0.86825 and an mAP50-95 of 0.62178. These findings suggest that BiFPN significantly strengthens the model's capability in feature extraction for multi-scale objects, while DSConv further optimizes computational efficiency and mitigates the impact of redundant features.

Compared with other models in the YOLO series, YOLOv8-BFDS achieves the best performance across all metrics. Notably, its leading advantage in the mAP50-95 metric is particularly significant, demonstrating that the improved model not only excels under the lenient evaluation criterion (mAP50) but also exhibits outstanding robustness and generalization capability under the more stringent evaluation criterion (mAP50-95). In contrast, YOLOv5 and YOLOv9 exhibit balanced precision and recall, but they fail to perform similarly in high-difficulty scenarios. Meanwhile, YOLOv7 and YOLOv6 show subpar performance across multiple metrics, revealing limitations in their architecture design and feature extraction capabilities on the current dataset.

The experimental results indicate that the improved YOLOv8-BFDS model demonstrates significant superiority in overall performance, outperforming all other models.

E. Comparison of Visualization Results with Other Methods

To provide a more intuitive comparison of the performance of different YOLO versions in underwater object detection, this study uses representative images from the JAMSTEC dataset, visualized across various scenarios. Bounding boxes of different colors represent different object categories detected by the models, with their corresponding category names and confidence scores. The comparison includes complex scenarios such as low-light environments, blurred imaging, and small object detection, highlighting the outstanding performance of the proposed YOLOv8-BFDS model in these situations.



YOLOv6 YOLOv7 YOLOv8 YOLOv9 YOLOv11 YOLOv8-BFDS (ours)

Fig. 2. Comparison of Object Detection Performance of Different YOLO Models in Four Typical Complex Underwater Scenarios

Authorized licensed use limited to: Ocean University of China. Downloaded on March 13,2026 at 11:00:08 UTC from IEEE Xplore. Restrictions apply.

As illustrated in Figure 2, bounding boxes of different colors represent various object categories detected by the models, with each box labeled with its corresponding category name and confidence score. It can be observed that the YOLOv8-BFDS model consistently achieves higher confidence scores compared to YOLOv6, YOLOv7, YOLOv9, YOLOv11, and the baseline YOLOv8 model across different scenarios. Furthermore, it demonstrates a higher detection rate for small objects and superior accuracy in multi-object and low-visibility conditions.

F. Ablation Study

To evaluate the impact of the optimization modules on YOLOv8's performance, we first added BiFPN and DSConv separately to the original model, then combined them to create YOLOv8-BFDS. All models were tested under identical conditions to evaluate each module's contribution to accuracy improvements. The ablation study results are as follows:

TABLE II. MODEL PERFORMANCE METRICS: YOLOV8-BFDS VS. YOLOV8 WITH BIFPN AND DSCONV

Model Configuration	Precision	Recall	mAP50
YOLOv8	0.95006	0.7985	0.86825
YOLOv8+BiFPN	0.97301	0.98314	0.98757
YOLOv8+DSConv	0.97051	0.98156	0.99089
YOLOv8-BFDS	**0.97011**	**0.98964**	**0.99149**

As illustrated in Table II, the training data for different groups in the ablation study are presented. The results show that the BiFPN module significantly enhances the model's detection accuracy and recall rate by optimizing feature fusion. Meanwhile, the DSConv module effectively reduces redundant computations and enhances feature extraction capabilities by incorporating depthwise separable convolutions. When combined, these two modules allow the YOLOv8-BFDS model to achieve new heights in both accuracy and recall rate compared to the baseline YOLOv8 model, demonstrating the effectiveness of the proposed approach. In particular, the substantial improvement in mAP50 (from 0.86825 to 0.99149) clearly highlights the model's outstanding performance in object detection tasks.

IV. CONCLUSION

In this paper, we address key challenges in seabed object detection, such as low lighting, blurred imaging, complex background interference, and small-object occlusion. We propose an optimized YOLOv8-based model, YOLOv8-BFDS, which enhances detection performance in underwater environments by integrating DCNv2, the EffectiveSE attention mechanism, and Concat_BiFPN.

Experimental results show that YOLOv8-BFDS outperforms the baseline YOLOv8 model across multiple key evaluation metrics, with notable improvements of 2.1% in precision and 19.1% in recall. It also demonstrates exceptional performance in mAP50 and mAP50-95, achieving 0.99149 and 0.83604, respectively. These results validate the advantages of the BiFPN module in multi-scale target feature extraction and the efficiency optimization provided by the DSConv module, which reduces interference from redundant features and enhances the model's performance in complex environments.

In summary, the YOLOv8-BFDS model demonstrates strong adaptability and precision in underwater object detection, outperforming other YOLO models in complex scenarios with challenging backgrounds, achieving higher detection accuracy. Future research will focus on optimizing the feature extraction network, exploring more efficient feature fusion and data augmentation techniques to enhance adaptability to multi-scale targets and complex environments. Additionally, combining data from multiple sensors (e.g., sonar, infrared) will further improve robustness and detection accuracy in intricate underwater scenarios. Efforts will also aim to enhance generalization and extend its application to other dynamic environments, advancing underwater object detection towards greater efficiency and precision.

REFERENCES

[1] Zhou Jianqun, Li Yang, Qin Hongmao, Dai Pengwen, Zhao Zilong, and Hu Manjiang. Sonar Image Generation by MFA-CycleGAN for Boosting Underwater Object Detection of AUVs[J]. IEEE Journal of Oceanic Engineering, 2024, 49(3): 905-919.

[2] Huang Yonghui, et al. Research on evaluation method of underwater image quality and performance of underwater structure defect detection model[J]. Engineering Structures, 2024, 306: 117797.

[3] Li Han. Research on Fish Target Detection System in Marine Pasture Based on Deep Learning[D]. Yantai: Yantai University, 2024.

[4] Zhang Fan. Research on Underwater Image Enhancement Techniques Oriented Towards Downstream Vision Tasks[D]. Nanning: Guangxi University, 2024.

[5] Han Tianbao, Wang Yue, Ren Shichang, Lv Xueqing. Underwater camouflage objects detection method for unmanned surface vessels[J]. Ship Science and Technology, 2024, 46(19): 85-91.

[6] Xie Shuang, Sun Hongwei. Tea-YOLOv8s: A Tea Bud Detection Model Based on Deep Learning and Computer Vision[J]. Sensors, 2024, 23: 6576-6599.

[7] Cao Wenwu, Li Taiqun, Wu Zhijia, Liu Jianzhuo. Multi-object tracking method based on YOLOv8 and quasi-dense similarity learning[J]. Chinese Journal of Lasers, 2025, 52(06): 0604003.

[8] Ye Jisong, Wu Yanjuan, Rong Wang. Based on the Optimization and Performance Evaluation of YOLOv8 Object Detection Model with Multi-backbone[C]// IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, 2024: 269-274.

[9] Zhang Enzhao, Li Lei, Wang Jianhua, Qin Jiewei, Liu Xuzhen, Hu Qiushi. Improved YOLOv8-based quality inspection method for ship plate welds[J]. Journal of Shaanxi University of Science and Technology, 2024, 42(6): 172-179.

[10] Yang Changchun, He Xuanxuan, Wang Rui, Zhu Shizhu, Yan Hao. Based on the improved YOLOv8 photovoltaic panel defect detection algorithm[J]. Electronic Measurement Technology, 2024, 47(23): 181-192.

Authorized licensed use limited to: Ocean University of China. Downloaded on March 13,2026 at 11:00:08 UTC from IEEE Xplore. Restrictions apply.