frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 November 2025
DOI 10.3389/fmars.2025.1673437

OPEN ACCESS

EDITED BY
Oscar Schofield,
Rutgers, The State University of New Jersey, United States

REVIEWED BY
Hao Wang,
Laoshan National Laboratory, China
Fickrie Muhammad,
Bandung Institute of Technology, Indonesia

*CORRESPONDENCE
Jianwei Huang
15559110766@163.com

RECEIVED 25 July 2025
ACCEPTED 24 October 2025
PUBLISHED 11 November 2025

CITATION
Huang Y, Huang J and Huang M (2025) A lightweight YOLO network for robotic underwater biological detection.
Front. Mar. Sci. 12:1673437.
doi: 10.3389/fmars.2025.1673437

A lightweight YOLO network for robotic underwater biological detection

Yanyu Huang[1], Jianwei Huang[2,3]* and Meihong Huang[1]

[1]College of Transportation and Navigation, Quanzhou Normal University, Quanzhou, China, [2]Naval University of Engineering, Wuhan, China, [3]Maritime College, Fujian Chuanzheng Communications College, Fuzhou, China

Introduction: Underwater image quality is commonly affected by problems such as insufficient illumination, extensive background noise, and target occlusion. Conventional biological detection methods suffer from the limitations of weak feature extraction, high computation, and low detection efficiency.
Methods: We propose an efficient and lightweight YOLO network for robots to realize high-precision underwater biological detection. Firstly, a backbone network based on hybrid dilated attention (HDA) is designed to expand the receptive field and focus on key features effectively. Secondly, a mixed aggregation star (MAS) network for the neck is constructed to enhance complex structural features and detailed textures of underwater organisms. Finally, the detection head is lightweighted using multi-scale content enhancement (MCE) modules to adaptively enhance key target channel information and suppress underwater noise.
Results: Compared to state-of-the-art target detection algorithms in underwater robots, our method achieves 85.7.% and 87.9% mAP@0.5 on the URPC2021 and the DUO datasets, respectively, with a model size of 5.19 M, a FLOP of 6.3 G, and a FPS of 16.54.
Discussion: The proposed method has excellent detection performance in underwater environments with low light, turbid water, and target occlusion.

KEYWORDS

underwater robot, biological detection, lightweight YOLO network, hybrid dilated attention, mixed aggregation star network, multi-scale content enhancement module

1 Introduction

Underwater target detection is a key technical means to realize the exploration of marine resources and promote the practice of marine engineering (Wang et al., 2024; Li et al., 2025). The underwater robot operates autonomously in different underwater environments through optical imaging devices and edge computing devices to accomplish the task of monitoring underwater biological targets, as shown in Figure 1. In particular, in the protection of rare and endangered species, underwater robots can utilize their advanced sensors and imaging technology to detect and count organisms, thus

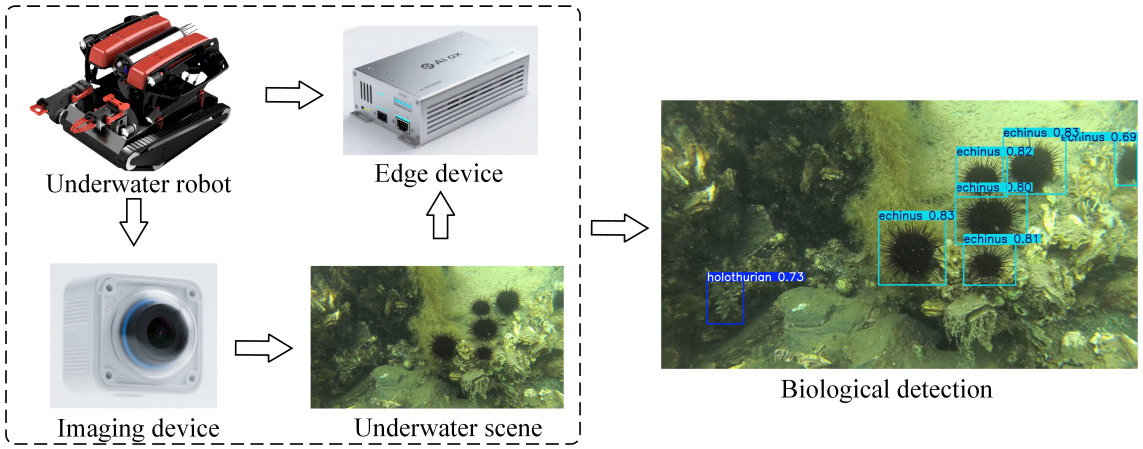

FIGURE 1
The process for underwater robotic biological detection.

grasping the changes in population size and providing a scientific basis for protection work. However, robots are in motion during underwater exploration, and it is difficult to obtain visually clear images, making accurate detection of underwater organisms still a significant technical bottleneck (Zheng et al., 2023).

Traditional underwater target detection techniques often follow the generalized framework of region proposal box generation, manual feature design, and classifier discrimination (Rout et al., 2024; Du et al., 2025). Due to the diversity of underwater environments, manually designed feature descriptors are difficult to adapt to organisms in different waters, depths, and lighting conditions, and cannot effectively extract multi-category organism features.

In the field of deep learning, neural networks excel in feature extraction and fitting capabilities (Abirami et al., 2025). The field of computer vision has also made some progress in underwater biomonitoring (Wang et al., 2024a; Wang et al., 2025). Currently, the two-stage algorithm improves the detection accuracy of underwater targets based on the framework of generating proposal regions to reclassify targets (Sun et al., 2025). The single-stage algorithm directly utilizes category regression to significantly reduce underwater target detection time, making the model more suitable for resource-limited platforms (Huang et al., 2024; Li et al., 2024; Wang et al., 2024).

However, in real underwater biological detection, robots face multiple challenges. (1) The underwater image quality is poor, suffering from color distortion, low contrast, and uneven lighting (Wang et al., 2024b). (2) Underwater organisms have diverse categories, large size changes, and obvious morphological differences, and the target may be obscured, tilted, or partially visible, which requires good robustness of the model. (3) The underwater environment is complex and variable, and the image background contains sand, rocks, corals, water plants, etc., with similar textures to the target, which increases the possibility of false detection. Therefore, further research on target detection algorithms applicable to underwater robots is needed to lighten the model structure and improve detection accuracy (Wang et al., 2024; Guo et al., 2025).

Given the computational limitations of underwater robotic platforms, we selected the YOLOv11 network as the foundational model for our algorithm. Although not the latest iteration of the YOLO architecture, YOLOv11 offers a mature deployment toolchain and exhibits low hardware requirements. Therefore, we innovate and optimize the YOLO11 network by combining the characteristics of underwater imaging and propose a lightweight YOLO network for robotic underwater biological detection. Specifically, in the backbone network, the HDA module is combined with the C3k2 module to enhance the model's ability to capture multi-scale features from underwater organisms. In the neck network, key features of low-contrast underwater organisms are enhanced by introducing the MAS network. In the probe head, the MCE module is used to accurately localize the contours of underwater organisms, effectively solving underwater detection challenges such as size variation, edge blurring, and background interference. The main contributions of this paper are as follows:

- An HDA module is constructed for extracting multi-scale target features from underwater images with color distortion, scale variation, and blurring degradation. The key features are effectively extracted by utilizing multi-scale null convolution fusion and hierarchical residual linkage strategies.
- To construct different-sized feature dependencies, the MAS network for feature fusion is designed. The key features of underwater organisms are adaptively enhanced through the star operation structure and a multi-branch fusion scheme.
- A lightweight MCE module for feature enhancement is proposed to improve the recognition ability of blurred underwater organisms. Through parallel multi-scale expansion of the convolutional structure and a shared parameter mechanism, the blurred features and multi-scale information of underwater organisms are dynamically enhanced.

- Extensive experiments on public datasets demonstrate that the proposed method has the advantages of excellent detection performance, lightweight models, and fast inference speed. Our method achieved 85.7% and 87.9% mAP@0.5 on the URPC2021 and DUO datasets, respectively. The model size is 5.19 MB, with 2.43 M parameters and 6.3 GFLOPs. The detection speed on a Jetson Nano 2GB device is 16.54 FPS.

The other sections are organized as follows: Section 2 summarizes related work. Section 3 introduces the innovative points of the proposed method. Section 4 describes the experimental setup and analyzes the experimental results. Section 5 presents the conclusions and future plans.

2 Related work

2.1 Underwater biological detection methods

Existing underwater target detection methods are mainly optimized based on generic target detection methods. These methods include the use of classical detection frameworks combined with techniques such as multi-scale feature fusion (Liu et al., 2025), feature weighting (Li et al., 2025), deformable convolution (Ouyang et al., 2024), and attention mechanisms (Tsai et al., 2025) to enhance detection.

Padmapriya et al. combined image enhancement techniques with deep convolutional neural networks, utilizing color correction, edge enhancement, and other operations to significantly enhance the expression of underwater target features, thereby addressing issues such as significant noise, low visibility, and uneven lighting conditions in underwater environments (Padmapriya et al., 2023). Liu et al. constructed a dual-path pyramid visual transformer feature extraction network, which cleverly utilized global features to enhance the differences between the foreground and background of images, thereby solving the problem of low accuracy in underwater fish detection (Liu et al., 2024). Li et al. proposed a self-supervised marine organism detection framework and designed an attention module specifically for underwater targets, thereby eliminating the dependence of underwater image data on annotation information (Li et al., 2024). However, the above methods have problems such as over-reliance on image preprocessing operations and high model complexity. It is difficult to be deployed in platforms with limited resources that cannot meet the requirements of real-time target detection in underwater scenes.

2.2 Lightweight methods

In practical underwater biological detection tasks, the challenge of limited computing and storage capacity of underwater detection equipment needs to be faced. The lightweight model can better adapt to the hardware conditions, quickly and accurately detect biological, meet the real-time requirements, and consume low energy, which can extend the range of the equipment and improve the detection efficiency. Therefore, many researchers have carried out research on the lightweighting of underwater biological detection models and achieved good results.

Li et al. proposed a lightweight underwater biological detection method by integrating a frequency attention mechanism with a dynamic convolution module, which improved feature extraction capabilities and solved the problems of large model parameters and high computational requirements (Li et al., 2025). Chen et al. proposed a lightweight aggregated underwater target detection network by constructing a multi-branch architecture combining convolutional and contextual attention, effectively solving the problem of target omission in biological target detection in complex underwater environments (Chen et al., 2024). Li et al. utilized reparameterization and global response normalization techniques to construct a feature enhancement and fusion network for underwater fuzzy object recognition, effectively reducing the impact of suspended particles in water on underwater target detection (Li and Cai, 2025). By using different network optimization techniques, the above method achieves lightweighting in the model structure, but there is still room for improvement in the detection of occluded biological and small-sized targets in complex underwater scenes.

2.3 Underwater target detection with YOLO

YOLO, as an end-to-end network, can reduce the error accumulation caused by complex underwater illumination and turbid water, effectively resist the interference of complex underwater environments (e.g., blurring, low-contrast, occlusion), and enhance the detection robustness. Therefore, the innovation and optimization of the YOLO network are of great significance to achieve high-precision biological detection in low-quality underwater images.

Zheng et al. innovatively introduced a reparameterized multi-scale fusion module and an aggregated distributed feature pyramid network into the YOLOv7 network, enabling the model to learn multi-scale features and thereby improving the detection performance for small underwater targets (Zheng and Yu, 2025). Liu et al. optimized the YOLOv8 network structure by using reparameterization techniques and spatial pyramid decomposition convolution to reduce target detail loss. And the introduction of cross-layer local attention in the detection header further reduces the computational cost and makes the model easier to deploy for edge computing devices (Liu et al., 2025). Ouyang et al. improved the YOLOv9 network using an attention block mechanism and an inflated large kernel network to enhance the local feature extraction and denoising capabilities, enabling the model to focus on underwater targets of different sizes (Ouyang and Li, 2025). Pan et al. constructed a lightweight marine biological detection model by improving YOLOv10. By introducing AKVanillaNet and DysnakeConv modules to enhance the target feature expression ability, and integrating Powerful-IOU loss function to optimize the

model training process, the performance of the model for target detection in underwater images with different lighting is improved (Pan et al., 2025). Therefore, by fully leveraging the rapid detection advantages of the YOLO network and innovatively optimizing its structure, the model's accuracy in detecting underwater targets of various sizes can be effectively improved, thereby meeting the practical needs of robotic underwater exploration.

3 Materials and methods

A lightweight YOLO network for underwater biological detection by robots is presented in this paper to achieve high-precision, rapid detection of small, occluded targets in low-light underwater images. The network structure of the proposed method is shown in Figure 2, where the spatial pyramid pooling fast (SPPF) and convolutional block with parallel spatial attention (C2PSA) modules are the original models of the YOLO11 network. Firstly, the HDA modules are introduced into the feature extraction network to capture multi-scale target features using different expansion rates, thus allowing the model to focus more on targets with large differences in size and shape in the underwater scene. Secondly, the MAS network in the neck region adaptively enhances feature expression and constructs feature information interaction channels at different scales through star computation and multi-branch feature fusion strategies. Finally, the adaptive feature enhancement property of the MCE module in the detection head is utilized to reduce the underwater noise interference and enhance the texture and edge features of low-contrast organisms to achieve the biological detection of complex underwater environments.

3.1 HDA module

In underwater multi-species, multi-size biological detection, the C3k2 module relies solely on a pure convolutional stacking structure, which cannot dynamically adjust the receptive field, making it difficult to effectively capture the features of underwater targets with different sizes. Moreover, the C3k2 module is difficult to establish global dependencies when extracting local features, and the global relationship between the target and its surroundings in the underwater scene is not taken into account, leading to limited detection accuracy in complex scenes. In this paper, we construct a C3k2_HDA feature extraction module based on hybrid dilated attention. The structure of the C3k2_HDA module is shown in Figure 3, where (A) denotes the C3k2 module, (B) denotes the C3k module, and (C) denotes the HDA module. The C3k2_HDA module is capable of effectively expanding the receptive field and focusing on key features by combining multi-scale hybrid cavity convolution (DConv), channel attention block (CAB), and

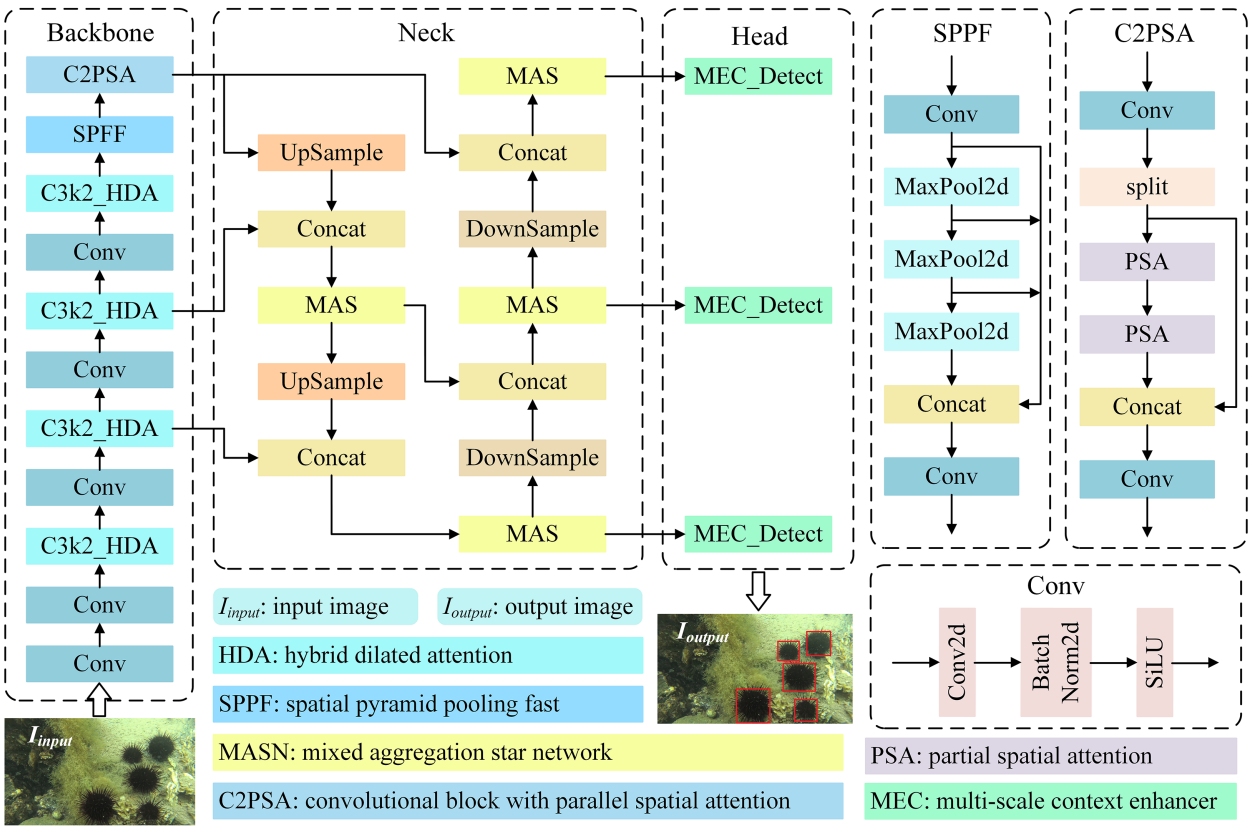

FIGURE 2
The network structure of the proposed method.

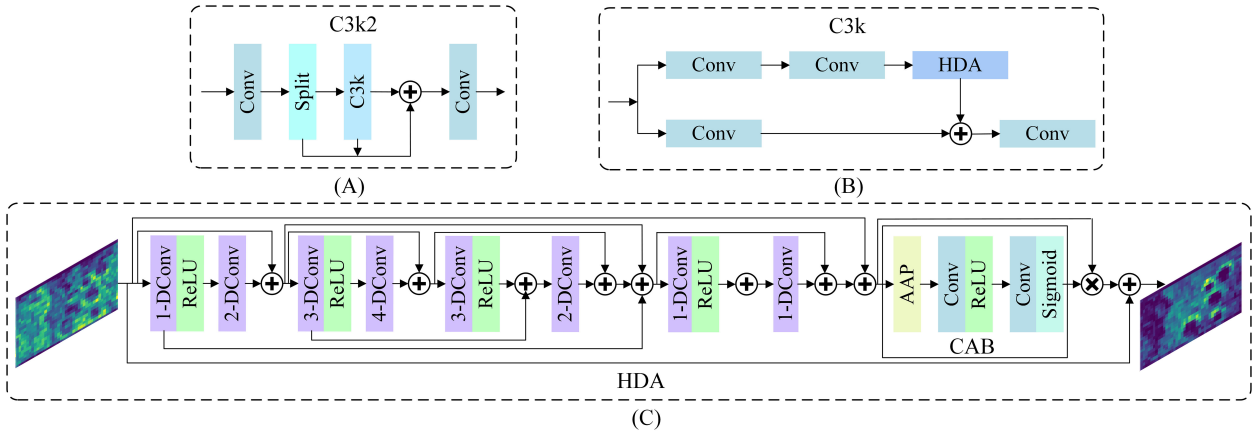

FIGURE 3
The structure of the C3k2_HDA model, where **(A–C)** represent the structures of the C3k2, C3k, and HDA models, respectively.

lightweight residual connectivity to extract multi-scale feature information of underwater organisms.

When given an input feature, the workflow of the C3k2_HDA module is as follows: Firstly, the feature channel is compressed using a convolution layer with a kernel of 1×1., which provides a finer representation of the feature for the subsequent convolutional operations, reducing the computational complexity, meanwhile highlighting the key features. Secondly, a fourth-order feature pyramid (3 × 3, 7 × 7, 13 × 13, 21 × 21) is constructed by gradually expanding the perceptual domain with a dilation rate of 1 to 4 through deep convolution. This ensures that the network can capture features at multiple scales, adapt to the diversity of shapes and sizes of underwater organisms, and is particularly useful for small targets that require a large receptive field. Thirdly, dynamic weighting between channels is achieved through a channel attention block (CAB), enabling the network to concentrate on key feature channels, suppress unimportant features, enhance the contrast between the target and the background, and reduce the impact of color distortion and noise interference in underwater images. The feature processing of CAB can be expressed as Equation 1.

$$CAB(y) = x \otimes \sigma(Conv_{1\times1}(ReLU(Conv_{1\times1}(AAP(x))))) \quad (1)$$

where x and y represent input features and output features, respectively. σ is the Sigmoid activation function, AAP is the adaptive average pooling function, and $\otimes$ is the channel-level multiplication.

Finally, by introducing a residual feature fusion mechanism and utilizing lightweight path connections to optimize detailed information, the integrity and accuracy of features are ensured, effectively reducing the impact of underwater image blurring and detail loss on biological detection. Therefore, the C3k2_HDA module helps the YOLO11 network extract as much feature information as possible from underwater images with color distortion and noise interference, thereby improving the accuracy, robustness, and adaptability of underwater biological detection.

3.2 MAS network

To enhance the blurred biometric expression of low-contrast underwater images, the MAS networks are introduced in the neck to replace the C3k2 module. The MAS network significantly optimizes underwater biological detection performance through a multi-branch feature fusion mechanism and a four-layer structure of star computation. The core idea is to utilize the 7×7 depth-separated convolution and two-way gating operation of the star module to expand the receptive field to suppress water turbidity and light noise, and to adaptively enhance the response of key regions such as biological contours and textures. The multi-branch structure then fuses features of different granularity (from edge details to semantic information) to enhance robustness to scale-variable, occluded targets. The design dramatically improves detection accuracy while maintaining computational efficiency. The structure of the MAS network is shown in Figure 4.

Given an input feature x with channel number c, the MAS network first performs feature decoupling by upscaling the feature channel to $2c$ through a convolution with a kernel 1×1, and subsequently splits it into four independent paths. The underwater target base features are extracted by a convolution with kernel of 1×1 in path 1, the spatial context information of the features is captured using a depth-separated convolution in path 2, the original feature is obtained by row-channel slicing operation in path 3, and a star operation is performed on the separated features in path 4 to get a richer and more expressive feature representation. In the star operation, a large receptive field is established by a 7×7 deep convolution to suppress underwater noise. Subsequently, x_1 and x_2 with a channel number of $3c$ are obtained using parallel convolution with kernel 1×1, and the ReLU6 activation function is applied to x_1, and then multiplied element-by-element with x_2 to realize feature filtering. The feature processing of star operation can be expressed as Equation 2.

$$y_1 = ReLU6(\boldsymbol{W}_1 * DWConv_{7\times7}(x)) \odot (\boldsymbol{W}_2 * DWConv_{7\times7}(x)) \quad (2)$$

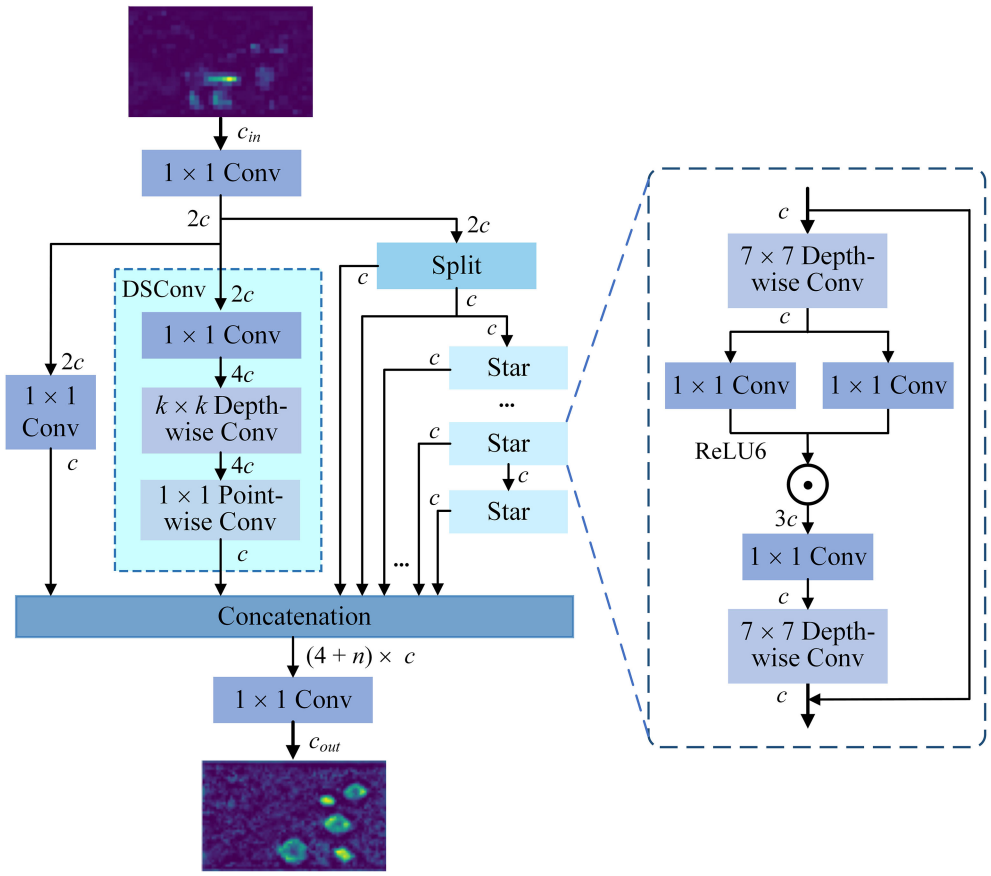

FIGURE 4
The structure of the MAS network.

where y_1 denotes the output after element-wise multiplication, $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ denote convolutional weights with kernel 1 × 1, and ⊙ denotes element-wise multiplication.

The channel compression and spatial information fusion are accomplished by the convolution with a kernel 1 × 1 and depth-wise convolution with a kernel 7 × 7. Combine the original input and output through the residual structure to obtain y_2, thus effectively ensuring the completeness of the features. The process can be expressed as Equation 3.

$$y_2 = x + DWConv_{7\times7}(Conv_{1\times1}(y_1)) \tag{3}$$

Finally, the initial four-path features are concatenated with the optimized features from the star module, fusing the multi-granularity information and outputting the enhanced feature map, which greatly strengthens the importance of pixels in the underwater target area. The process can be expressed as Equation 4.

$$y_{MAS} = Conv_{1\times1}\left(y_{p1} \oplus y_{p2} \oplus y_{p3} \oplus y_{p4} \oplus \sum_{k=1}^{n} y_{star}\right) \tag{4}$$

where y_{MAS} denotes the output of the MAS network, y_{p1}, y_{p2}, y_{p3}, and y_{p4} denote the outputs of the four paths, respectively, y_{star} denotes the output of the star module, and ⊕ denotes the concatenation operation.

3.3 MCE module

To address the limitations of the YOLO11 network detector head in dealing with small-sized, irregular, and densely occluded targets underwater, we propose a multi-scale content enhancement module and combine it with a shared convolution detector to form a lightweight and highly efficient LSCD_MEC detector as shown in Figure 5. The core idea of the MEC module is to use three-way dilated convolutions (dilation = 2/4/6) to capture different receptive fields of contextual information, thereby constructing feature data dependencies at different scales and achieving data fusion. The channel attention mechanism can further dynamically enhance biometric features and suppress interference from murky water backgrounds. Finally, the original feature is introduced in the fusion stage to avoid the loss of high-frequency detail information. Therefore, the feature sharing and lightweight design of the LSCD_MEC detector optimizes the model structure, reduces the number of parameters, and improves the stability of multi-scale biological detection.

When three different scale features from the neck are input to the head, the LSCD_MEC module first performs hierarchical preprocessing of the features using a group normalization operation to extract the initial features of the underwater target. Subsequently, in the MCE module, the features are divided into four

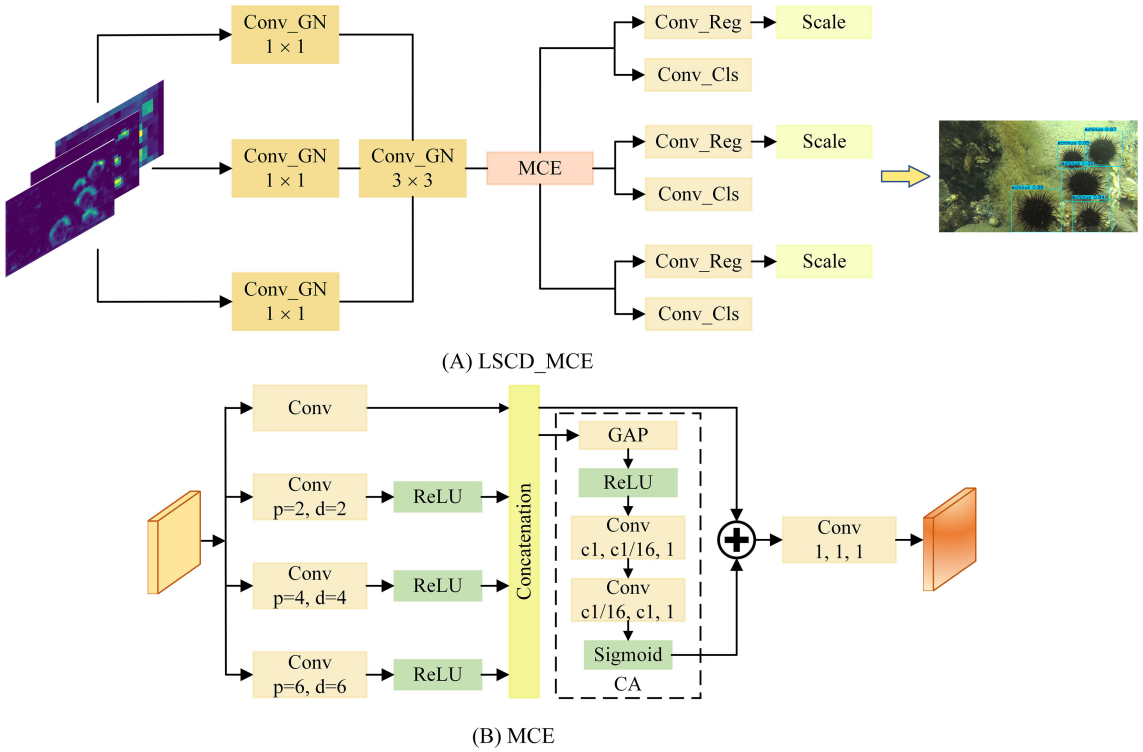

FIGURE 5
The structure of the LSCD_MCE module.

equal parts along the channel dimension. Local details, medium sensory fields, and large-scale contexts are captured using depth-wise convolution with three different expansion rates. The features are fused and transmitted to the channel attention (CA) module to improve the model's focus on important features of underwater organisms. The channel attention weighting process can be expressed as Equation 5.

$$y_{CA} = \sigma(W_2(ReLU(W_1(GAP(x))))) \quad (5)$$

where y_{CA} and GAP represent the output of the CA module and the global average pooling operation, respectively.

Finally, two convolutions with a kernel of 1 × 1 are used to construct a regression branch and a classification branch, outputting the bounding box parameters and the category probabilities, and completing the accurate detection of underwater targets.

4 Experiment preparation

4.1 Experimental setting and parameters

All the ablation experiments and comparison tests in this paper were conducted under a device equipped with a deep learning framework. The model training parameters and device's detailed configuration are shown in Table 1.

4.2 Datasets

In this paper, the publicly available underwater scene biological image dataset, the underwater robot programming contest 2021 (URPC2021), and detecting underwater objects (DUO) are utilized to test the performance of the proposed method in underwater biological detection. The two datasets involve consistent categories, namely holothurian, echinus, scallop, and starfish. Both datasets contain complex underwater scenes containing low light, low contrast, target occlusion, size inconsistency, etc., which can comprehensively evaluate the practicality of the optimized YOLO11 model for biological detection in complex underwater scenes.

The UPRC2021 dataset contains 7,600 annotated underwater images, which are divided into training and testing sets at a ratio of 8:2 during training. The original annotations are in Pascal VOC standard XML format, which is converted to YOLO format before the experiment.

The DUO dataset integrates the datasets from the URPC Challenge over the years, removes the duplicates, and re-labels the erroneous labels. The DUO dataset contains a total of 7,782 accurately annotated images, with a training set to test set ratio of 8:2. The images of the DUO dataset present excellent bias, low contrast, and uneven illumination, blurring and high noise and other typical underwater image characteristics, which pose certain challenges for accurate detection of different aquaculture organisms,

TABLE 1 Detailed information on device configuration and model training parameters.

Experimental setup	Parameters	Value
Experimental device	operating system	Windows 11
	CPU	13th Intel® Core™ i9-13900KF 3.00 GHz
	GPU	Nvidia GeForce GTX 3090 with CUDA 12.4 and cuDNN 8.9
	RAM	128GB
	framework	PyTorch 2.5.0, Python 3.8.11
Training parameters	learning rate	0.01
	momentum	0.937
	epochs	500
	batch size	64
	images size	640 × 640
	close mosaic	10
	weight decay	0.0005
	device	1
	optimizer	SGD
	automatic mixed precision	true
	degrees	60
	scale	0.5
	shear	60
	perspective	0.001

meanwhile largely reflecting the problems faced by real marine environment detection targets, and providing a unified benchmark for the evaluation of underwater target detection algorithms.

4.3 Evaluation metrics

In this paper, the following metrics are introduced: accuracy, recall rate, mAP50, model parameters, model size, and GFLOPs to objectively evaluate the performance of various methods on biological detection in complex underwater scenes (Girshick et al., 2014).

Precision reflects the reliability of the model's prediction of positive samples of underwater organisms, while recall indicates the probability that the model correctly identifies positive samples of underwater organisms. The calculation of precision and recall can be expressed as Equations 6, 7.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

where TP denotes a predicted positive sample of underwater organisms and an actual positive sample of underwater organisms, FP denotes a predicted positive sample of underwater organisms and an actual negative sample of underwater organisms, and FN denotes a predicted negative sample of underwater organisms but an actual positive sample of underwater organisms.

The average precision (AP) is used to comprehensively evaluate the accuracy of a model at different recall levels. The mean average precision (mAP) can measure the overall performance of a model for all categories. The calculations for AP and mAP can be expressed as Equations 8, 9.

$$AP = \int_0^1 Precision(Recall)\mathrm{d}(Recall) \tag{8}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{9}$$

where N denotes the number of categories, $N = 4$. mAP@0.5 and mAP@0.95 represent the mAP values when IoU=0.5 and IoU=0.95, respectively.

Model parameters are used to measure model scale and complexity. Model size is used to determine the ease of deployment and resource consumption of the model for underwater detection. Floating point operations (FLOPs) are directly related to model inference latency and hardware computing power requirements.

5 Experimental results and analysis

5.1 Ablation experiments

To quantitatively demonstrate the improvement in underwater biological detection performance achieved by each innovation module, ablation experiments are conducted on the URPC2021 dataset, with the results shown in Table 2.

The YOLO11 model, without any innovation point, achieves 82.4% mAP@0.5 and 49.8% mAP@0.95, with a model size of 5.25MB, FLOPs of 6.4G, and a parameter of 2.59×10^6. After inserting the HDA module into the backbone network, mAP@0.5 and mAP@0.95 improve by 1.2% and 1.1%, but the model size, FLOPs, and parameters increase by 0.09MB, 0.3G, and 0.02×10^6. After introducing the MAS network in the neck network, mAP@0.5 and mAP@0.95 increased to 84.6% and 51.8%, respectively, with the model size of 5.42 MB, FLOPs of 6.9, and parameters of 2.69×10^6. After using the MCE module in the head, mAP@0.5 and mAP@0.95 reach 85.7% and 52.9%, respectively, and the model size is only 5.19 M, the FLOPs are 6.3 G, and the parameters are 2.43 M. Therefore, the introduction of the HDA module and MAS network can effectively improve the detection performance of underwater

TABLE 2 Effectiveness of innovation points for detection.

YOLO11	HDA	MAS	MCE	mAP@0.5/%	mAP@0.95/%	Model size/MB	FLOPs/G	Parameter/10^6
✓				82.4%	0.498	5.25	6.4	2.59
✓	✓			83.6%	0.509	5.34	6.7	2.61
✓	✓	✓		84.6%	0.518	5.42	6.9	2.69
✓	✓	✓	✓	**85.7%**	**0.529**	**5.19**	**6.3**	**2.43**

The bold text indicates the best scores achieved in the experiments.

organisms, while the introduction of the MCE module significantly reduces the memory footprint and computational complexity. The proposed method is more favorable for deployment in equipment with limited resources.

To visualize the effect of innovation points in underwater biological detection, gradient-weighted class-activation mapping is introduced to generate heatmaps for the YOLO11 model and each innovation point. The results are shown in Figure 6, where (a) is the original input image, (b) is the heat map of YOLO11, (c)-(e) are the heat maps generated by introducing the HDA, MAS, and MCE modules, respectively, and (f) is the ground-truth image. The red areas in the heat map indicate that the model contributes more to the detection of underwater organisms. The three groups of images represent color distortion images, normal underwater images, and low-light images, respectively. In images with color distortion, YOLO11 only focuses on objects with a large target scale, missing small-sized targets. On the contrary, after introducing the HDA, MAS, and MCE modules, the model's attention gradually covered small targets, indicating that the optimized YOLO11 can effectively capture underwater biological features of different scales. In normal underwater images, YOLO11 only focuses on targets with obvious features and omits occluded targets. With the HDA, MAS, and MCE modules, the model not only effectively detects the occluded targets but also can reduce the background interference and successfully detects the starfish with a similar color to the seabed. In low-light images, YOLO11 cannot effectively distinguish the area where the target is located and suffers from detection errors. The proposed method can effectively find out the targets hidden in a low-light environment and realize the accurate detection of organisms.

5.2 Comparative experiments

To further validate the proposed method's ability to accurately identify underwater organisms in complex underwater environments, Faster RCNN, YOLOv5, and other methods are introduced to compare the performance of URPC2021 and DUO datasets. The performance index scores are shown in Tables 3 and 4, respectively.

Table 3 shows that among the various methods, Faster RCNN scored the lowest in the evaluation metrics, with the largest model size, FLOPs, and number of parameters, indicating the highest computational complexity and the greatest resource requirements for underwater organism detection. YOLOv5s and RTD-YOLOv5 achieve good results in precision, recall, and mAP metrics, but do not have an advantage in model size. YOLOv7 network scores poorly in all evaluation metrics, which not only has low underwater target detection accuracy, but also consumes more memory for

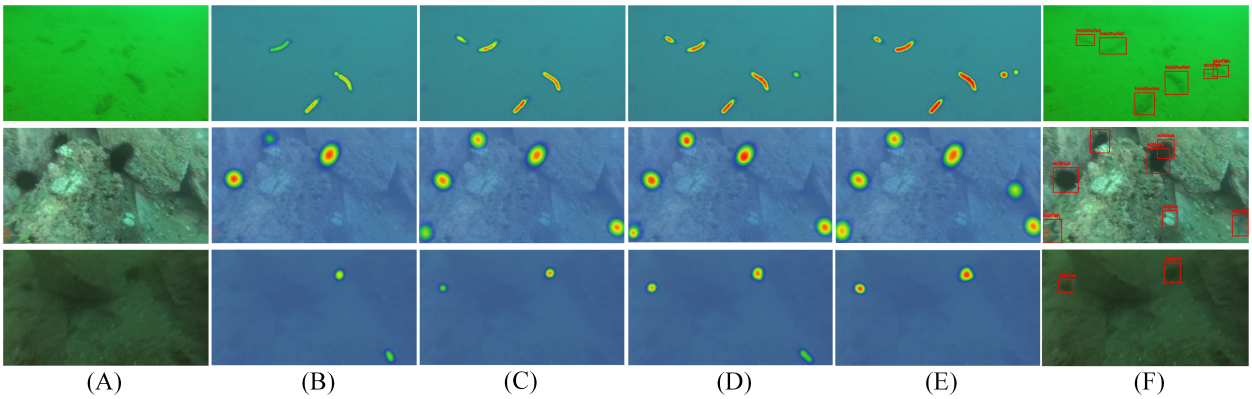

FIGURE 6
Heat maps of YOLO11 model and innovation points, where **(A)** shows the original image, **(B)** displays the YOLO11 heatmap, and **(C–E)** present heatmaps generated after integrating the HDA, MAS, and MCE modules into YOLO11, respectively. **(F)** represents the ground-truth image.

TABLE 3 The detection scores of each method on the URPC2021 dataset.

Method	Precision/%	Recall/%	mAP@0.5/%	mAP@0.95/%	Model Size/MB	FLOPs/G	Parameter/10^6
Faster RCNN (Ren et al., 2016)	75.2	64.6	74.3	41.5	41.3	210	1347
YOLOv5s (Zhu et al., 2021)	83.1	76	82.4	46.5	16.1	16.5	7.2
RTD-YOLOv5 (Yuan et al., 2024)	84.3	72.8	82.4	45.9	14.6	12.3	5.9
YOLOv7 (Wang et al., 2023)	81.6	75.2	82.4	47.2	72.1	103.3	36.5
YOLOv8	82.6	76.3	82.3	48.9	6.2	8.7	3.2
YOLOv8-LA (Qu et al., 2024)	84.9	76.8	84.7	50.2	5.9	7.5	**2.4**
YOLOv10s	85.2	77.9	83.7	51.2	7.5	24.5	8.0
YOLO11	84.4	76.6	82.4	49.8	5.3	6.4	2.6
Ours	**85.9**	**78.2**	**85.7**	**52.9**	**5.19**	**6.3**	2.43

The bold text indicates the best scores achieved in the experiments.

model deployment. The YOLOv8 and YOLOv8-LA methods achieved excellent results across all evaluation metrics, not only accurately detecting biological objects in complex underwater environments but also featuring compact model sizes and low computational complexity, making them suitable for underwater robots. Compared to other YOLO methods, YOLOv10s has greater computational requirements. The method proposed in this paper outperforms the YOLOv11 network by 1.5% and 1.6% in precision and recall, respectively. Furthermore, mAP@0.5 and mAP@0.95 achieve the highest scores, with the smallest model size, number of parameters, and FLPOs. Therefore, our method can accurately detect small organisms that are obscured in complex underwater scenes, effectively solving the problem of target detection misses. The model is small in size and low in computational complexity, enabling it to efficiently perform real-time underwater organism detection tasks.

In the DUO dataset test, the Cascade R-CNN and Boosting RCNN methods achieved the smallest mAP scores, and their large model sizes and high computational complexity are not conducive to efficient underwater target detection. The Deformable DET and RTMDet models, although they achieved better scores in detection performance, are limited by the resources consumed by the models. Compared with YOLO10s, YOLOv7, and YOLO11, their model sizes and FLOPs require more resources. YOLOv5s and YOLOv8 models achieved 0.834 and 0.851 mAP@0.5, which showed excellent detection performance, but their model sizes and FLOPs

TABLE 4 The detection scores of each method on the DUO dataset.

Method	mAP@0.5/%	mAP@0.95/%	Model Size/MB	FLOPs/G	Parameter/10^6
Cascade R-CNN (Cai and Vasconcelos, 2018)	82.1	61.2	44.5	91.1	68.9
Deformable DET (Zhu et al., 2020)	84.4	63.7	44.7	173	40.0
Boosting R-CNN (Song et al., 2023)	78.5	63.5	125.1	53.2	43.6
RTMDet (Lyu et al., 2022)	83.2	63.8	125.5	39.1	24.7
YOLOv5s (Zhu et al., 2021)	83.4	62.1	16.1	16.5	7.2
YOLOv7 (Wang et al., 2023)	82.6	61.4	18.6	103.3	5.9
YOLOv8	85.1	65	39.4	28.4	11.1
YOLOv10s (Wang et al., 2024)	84.6	64.8	7.5	24.5	8.0
YOLO11	80.2	60.4	5.2	20.4	6.4
RG-YOLO (Zheng and Yu, 2025)	86.1	65.7	7.4	31.1	10.1
Ours	**87.9**	**67.3**	**5.19**	**6.3**	**2.43**

The bold text indicates the best scores achieved in the experiments.

10.3389/fmars.2025.1673437

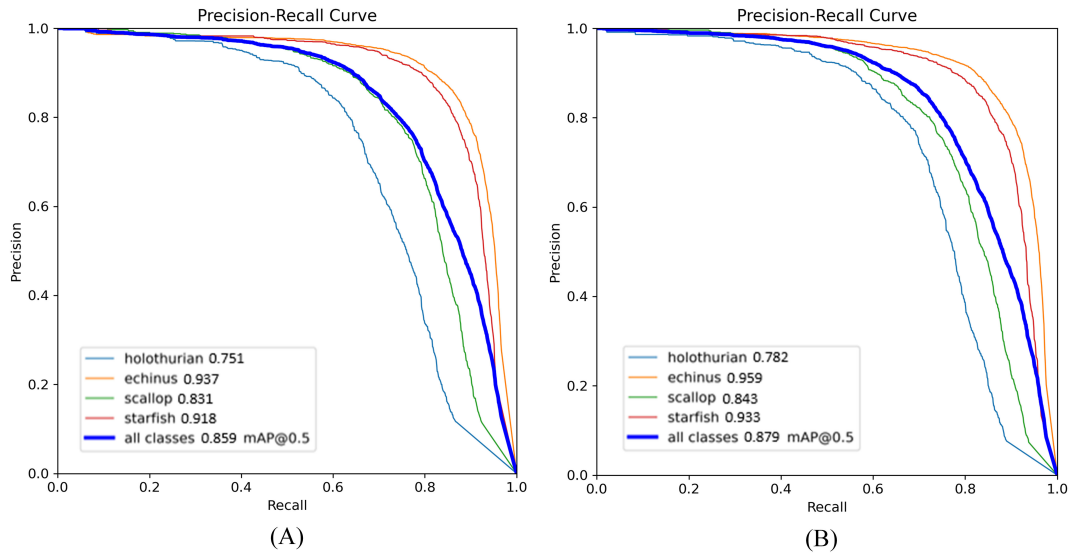

FIGURE 7
Detection scores for each category using the proposed method, where **(A)** corresponds to the URPC2021 dataset and **(B)** corresponds to the DUO dataset.

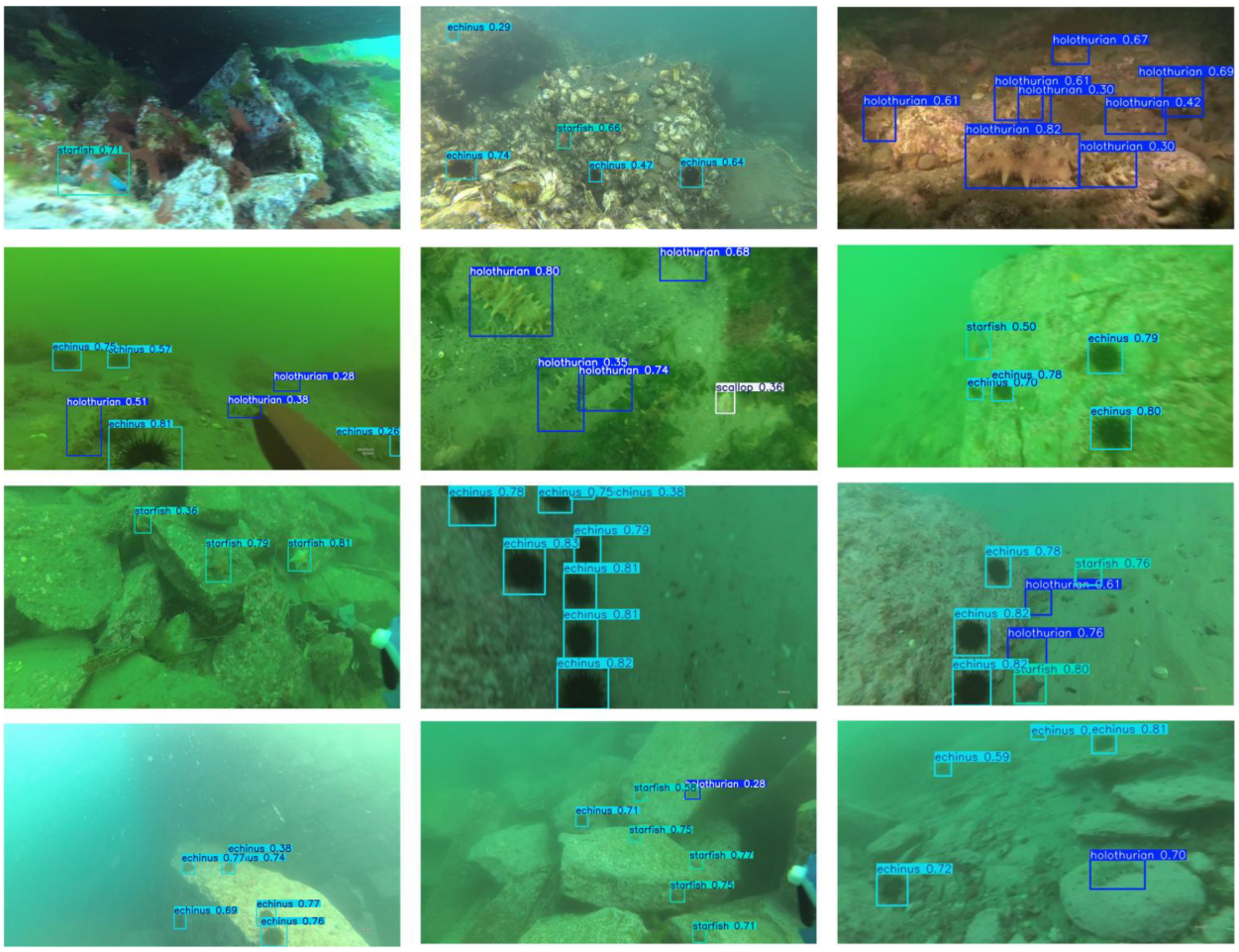

FIGURE 8
The detection results of the proposed method on URPC2021 dataset.

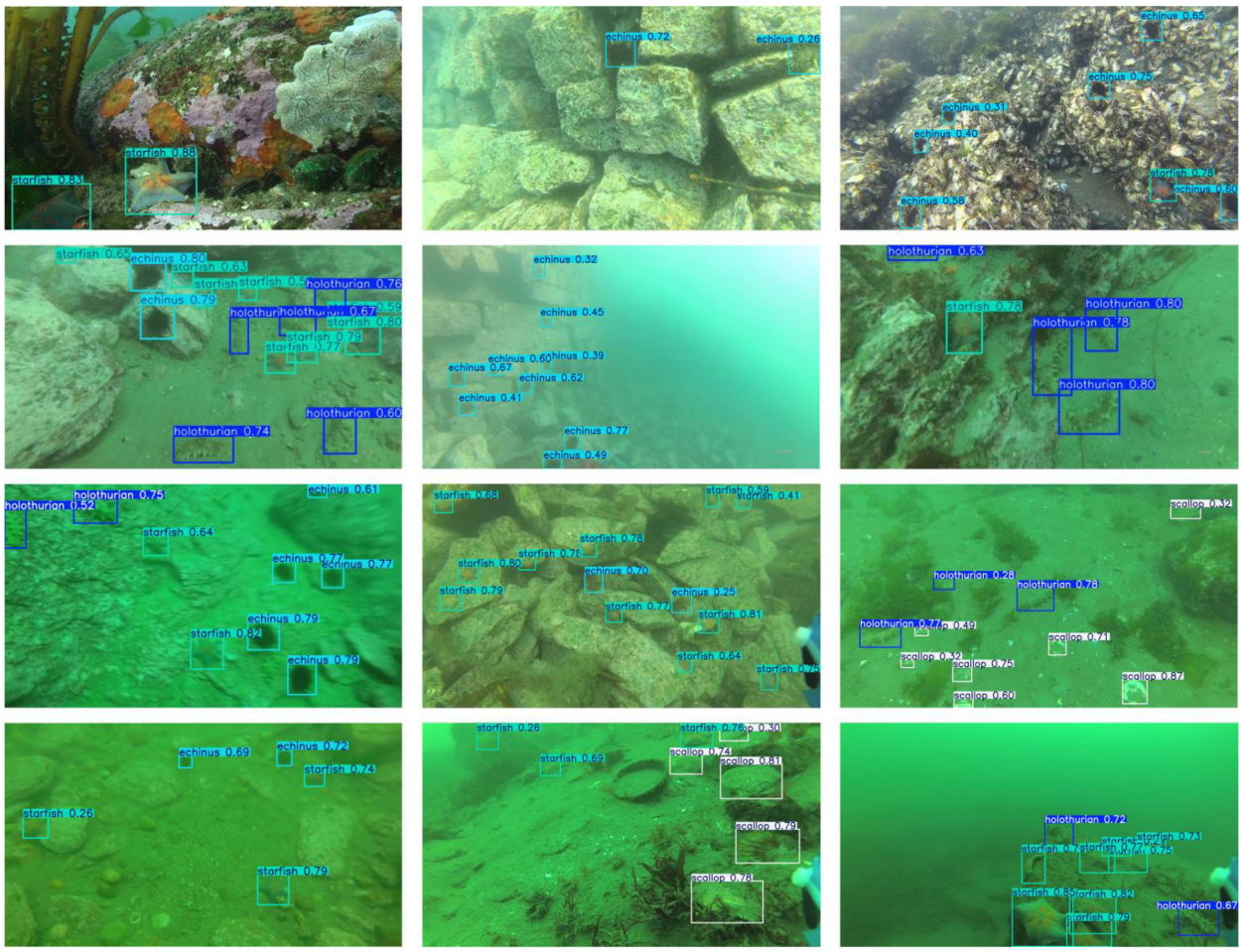

FIGURE 9
The detection results of the proposed method on the DUO dataset.

required more resources than YOLO10s and YOLO11. RG-YOLO improved the YOLO method and achieved significant improvement in detection performance, but there is still much room for improvement in the model's computational complexity. The method in this paper achieved 87.9% and 67.3% on mAP@0.5 and mAP@0.95, with a model size of 5.19 M, FLOPS of 6.3 G, and parameter number of 2.43 × 106, which not only demonstrated the most excellent detection performance, but also low computational complexity and low parameter number, which provided an effective technological solution for real-time and efficient biological detection in complex underwater scenes.

The detection mAP scores of the proposed method for each category are shown in Figure 7, (a) for the URPC2021 dataset and (b) for the DUO dataset. The proposed model performed best in echinus, which achieves the highest score among the four categories, but has weakest detection performance for holothurian due to the fact that holothurian is similar to the seabed in imaging. Overall, the proposed method presents excellent underwater biological detection performance.

Figures 8 and 9 show the results of biological detection using our method on the URPC2021 and DUO datasets. In the URPC2021 and DUO datasets, the proposed method shows excellent detection accuracy in underwater environments with multi-sized targets, low light, color distortion, blurring, and noise. In particular, the proposed method also accurately detects imaging-obscured mimicry biological detection, which is able to adapt to complex and changing underwater environments. In addition, we deploy the YOLO series methods on Jetson Nano 2GB devices for testing, with results shown in Table 5. Our method achieved 16.54 FPS in the test of the Jetson Nano 2GB device, demonstrating the feasibility of the method for underwater robotic deployment.

TABLE 5 The FPS of the YOLO network on Jetson Nano 2GB devices.

Method	FPS
YOLOv5s (Zhu et al., 2021)	8.39
YOLOv7	12.61
YOLOv8	9.33
YOLOv10s (Wang et al., 2024)	11.42
YOLO11	14.92
Ours	**16.54**

The bold text indicates the best scores achieved in the experiments.

6 Conclusions

A lightweight YOLO network for robotic underwater biological detection is proposed, aiming to help underwater robots efficiently accomplish underwater resource exploration tasks. The backbone network based on the HDA module effectively suppresses underwater image noise interference and improves the model's attention to targets in low-light environments. The MAS network is designed to achieve feature dynamic optimization and efficient multi-scale information interactive fusion, which solves the problem that target detection is easy to miss under the occlusion of underwater scenes. A MCE module is proposed to adaptively enhance key information of multiple-scale features, thereby improving the detection performance of fuzzy targets. Finally, the proposed method obtained the highest detection scores in both URPC2021 and DUO datasets in the comparison experiments. Moreover, the feasibility of the proposed method for underwater robotic deployment was verified in a Jetson Nano 2GB device. Therefore, our method demonstrates outstanding detection performance in underwater biological detection, meeting the requirements of actual underwater resource exploration projects for effectiveness and real-time performance.

In subsequent work, we will apply this method to underwater robotic systems, cross-validate the algorithm using stereoscopic camera systems underwater, and investigate the impact of different water body parameters on underwater detection performance. Through additional practical underwater exploration missions, we will continuously optimize the model's performance and practicality.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

YH: Conceptualization, Funding acquisition, Methodology, Software, Writing – original draft, Writing – review & editing. JH: Methodology, Software, Writing – original draft, Writing – review & editing. MH: Software, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported in part by the Fujian Provincial Department of Education Project under Grant JAT190531, in part by the Quanzhou Normal University Students' Innovation and Entrepreneurship Training Program Funded Project under Grant S202410399061X, in part by the Xiamen Marine and Fishery Development Special Fund Project under Grant 21CZB013H115, and in part by the Xiamen Key Laboratory of Marine Intelligent Terminal Development and Application under Grant B18208.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Abirami, G., Nagadevi, S., Jayaseeli, J. D. D., Rao, T. P., Patibandla, R. S. M. L., Aluvalu, R., et al. (2025). An integration of ensemble deep learning with hybrid optimization approaches for effective underwater object detection and classification model. *Sci. REP-UK.* 15, 10902. doi: 10.1038/s41598-025-95596-5

Cai, Z., and Vasconcelos, N. (2018). "Cascade R-CNN: delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).* (Salt Lake, Utah, USA: IEEE). 6154–6162. doi: 10.1109/CVPR.2018.00644

Chen, X., Fan, C., Shi, J., Chen, X., and Wang, H. (2024). Underwater-MLA: lightweight aggregated underwater object detection network based on multi-branches for embedded deployment. *MEAS. Sci. Technol.* 36, 016192. doi: 10.1088/1361-6501/ad9b42

Du, X., Wen, Y., Yan, J., Zhang, Y., Luo, X., and Guan, X. (2025). Multi-target detection in underwater sensor networks based on bayesian deep learning. *IEEE T. Netw. Sci. ENG.* 12, 1581–1596. doi: 10.1109/TNSE.2025.3535572

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR).* (Seattle, Washington, USA: IEEE). 580–587.

Guo, L., Liu, X., Ye, D., He, X., Xia, J., and Song, W. (2025). Underwater object detection algorithm integrating image enhancement and deformable convolution. *Ecol. INFORM.* 89, 103185. doi: 10.1016/j.ecoinf.2025.103185

Huang, J., Wang, K., Hou, Y., and Huang, J. (2024). LW-YOLO11: A lightweight arbitrary-oriented ship detection method based on improved YOLO11. *SENSORS-BASEL* 25, 65. doi: 10.3390/s25010065

Li, B., and Cai, L. (2025). REFNet: reparameterized feature enhancement and fusion network for underwater blur target recognition. *ROBOTICA*. 43, 1867–1884. doi: 10.1017/S0263574725000475

Li, N., Ding, B., Yang, G., Ni, S., and Wang, N. (2025). Lightweight LUW-DETR for efficient underwater benthic organism detection. *Visual Comput.*, 41, 1–16. doi: 10.1007/s00371-025-03921-w

Li, M., Li, J., and Feng, H. (2024). Detection and recognition of underwater acoustic communication signal under ocean background noise. *IEEE Access* 12, 149432–149446. doi: 10.1109/ACCESS.2024.3476494

Li, S., Wang, Z., Dai, R., Wang, Y., Zhong, F., and Liu, Y. (2025). Efficient underwater object detection with enhanced feature extraction and fusion. *IEEE T. Ind. INFORM.* 21, 4904–4914. doi: 10.1109/TII.2025.3547007

Li, J., Yang, W., Qiao, S., Gu, Z., Zheng, B., and Zheng, H. (2024). Self-supervised marine organism detection from underwater images. *IEEE J. OCEANIC ENG.* 50, 120–135. doi: 10.1109/JOE.2024.3455565

Li, J., Zhao, L., Li, H., Xue, X., and Liu, H. (2025). MixRformer: dual-branch network for underwater image enhancement in wavelet domain. *SENSORS-BASEL* 25, 3302. doi: 10.3390/s25113302

Liu, Y., An, D., Ren, Y., Zhao, J., Zhang, C., and Chen, J. (2024). DP-FishNet: dual-path pyramid vision transformer-based underwater fish detection network. *Expert Syst. Appl.* 238, 122018. doi: 10.1016/j.eswa.2023.122018

Liu, M., Wu, Y., Li, R., and Lin, C. (2025). LFN-YOLO: precision underwater small object detection via a lightweight reparameterized approach. *Front. Mar. Sci.*, 11, 111513740–1513740. doi: 10.3389/fmars.2024.1513740

Liu, C., Yao, H., Qiu, W., Cui, H., Fang, Y., and Xu, A. (2025). Multi-scale feature map fusion encoding for underwater object segmentation. *Appl. Intell.* 55, 1–17. doi: 10.1007/s10489-024-05971-4

Lyu, C., Zhang, W., Huang, H., Zhou, Y., Wang, Y., and Liu, Y. (2022). RTMDet: an empirical study of designing real-time object detectors. *arxiv preprint arxiv:2212.07784*. doi: 10.48550/arXiv.2212.07784

Ouyang, J., and Li, Y. (2025). Enhanced underwater object detection via attention mechanism and dilated large-kernel networks. *Visual Comput.*, 41, 1–23. doi: 10.1007/s00371-025-03870-4

Ouyang, W., Wei, Y., and Liu, G. (2024). A lightweight object detector with deformable upsampling for marine organism detection. *IEEE T. INSTRUM. MEAS.* 73, 1–9. doi: 10.1109/TIM.2024.3385846

Padmapriya, S., Umamageswari, A., Deepa, S., and Banu, J. (2023). A novel deep learning based underwater image de-noising and detecting suspicious object. *J. Intell. FUZZY Syst.* 45, 7129–7144. doi: 10.3233/JIFS-234002

Pan, W., Chen, J., Lv, B., and Peng, L. (2025). Lightweight marine biodetection model based on improved YOLOv10. *ALEX. ENG. J.* 119, 379–390. doi: 10.1016/j.aej.2025.01.077

Qu, S., Cui, C., Duan, J., Liu, Y., and Pang, Z. (2024). Underwater small target detection under yolov8-la model. *Sci. REP-UK.* 2024, 14, 16108. doi: 10.1038/s41598-024-66950-w

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE T. Pattern Anal.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rout, D. K., Kapoor, M., Subudhi, B. N., Thangaraj, V., Jakhetiya, V., and Bansal, A. (2024). Underwater visual surveillance: a comprehensive survey. *OCEAN ENG.* 309, 118367. doi: 10.1016/j.oceaneng.2024.118367

Song, P., Li, P., Dai, L., Wang, T., and Chen, Z. (2023). Boosting R-CNN: reweighting r-cnn samples by rpn's error for underwater object detection. *NEUROCOMPUTING* 530, 150–164. doi: 10.1016/j.neucom.2023.01.088

Sun, H., Yue, A., Wu, W., and Yang, H. (2025). Enhanced marine fish small sample image recognition with rvfl in Faster R-CNN model. *Aquaculture* 595, 741516. doi: 10.1016/j.aquaculture.2024.741516

Tsai, Y., Tsai, C., and Huang, J. (2025). Multi-scale detection of underwater objects using attention mechanisms and normalized wasserstein distance loss. *J. SUPERCOMPUT.* 81, 1–33. doi: 10.1007/s11227-025-07251-5

Wang, C., Bochkovskiy, A., and Liao, H. Y. M. (2023). "YOLOv7: trainable bag-0f-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. (Vancouver, British Columbia, Canada: IEEE). 7464–7475. doi: 10.1109/CVPR52729.2023.00715

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., and Han, J. (2024). YOLOv10: real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* 37, 107984–108011. doi: 10.48550/arXiv.2405.14458

Wang, Z., Ruan, Z., and Chen, C. (2024). DyFish-DETR: underwater fish image recognition based on detection transformer. *J. Mar. Sci. ENG.* 12, 864. doi: 10.3390/jmse12060864

Wang, H., Sun, X., Chang, L., Li, H., Zhang, W., Frery, A. C., et al. (2024a). INSPIRATION: A reinforcement learning-based human visual perception-driven image enhancement paradigm for underwater scenes. *Eng. Appl. Artif. Intelligence.* 133, 108411. doi: 10.1016/j.engappai.2024.108411

Wang, W., Sun, Y. F., Gao, W., Xu, W., Zhang, Y., and Huang, D. (2024). Quantitative detection algorithm for deep-sea megabenthic organisms based on improved YOLOv5. *Front. Mar. Sci.* 11, 1301024. doi: 10.1007/s00530-025-01846-x

Wang, H., Sun, X., and Ren, P. (2024b). Underwater color disparities: cues for enhancing underwater images toward natural color consistencies. *IEEE Trans. Circuits Syst. Video Technol.* 34, 738–753. doi: 10.1109/TCSVT.2023.3289566

Wang, M., Zhang, K., Wei, H., Chen, W., and Zhao, T. (2024). Underwater image quality optimization: researches, challenges, and future trends. *IMAGE Vision Comput.*, 146, 104995. doi: 10.1016/j.imavis.2024.104995

Wang, H., Zhang, W., Xu, Y., Li, H., and Ren, P. (2025). WaterCycleDiffusion: Visual-textual fusion empowered underwater image enhancement. *Inf. Fusion*, 127, 103693. doi: 10.1016/j.inffus.2025.103693

Yuan, S., Luo, X., and Xu, R. (2024). "Underwater robot target detection based on improved YOLOv5 network," in *2024 12th International Conference on Intelligent Control and Information Processing (ICICIP)*. (Nanjing, Jiangsu, China: IEEE). 33–38. doi: 10.1109/ICICIP60808.2024.10477835

Zheng, Z., and Yu, W. (2025). RG-YOLO: multi-scale feature learning for underwater target detection. *MULTIMEDIA Syst.* 31, 26. doi: 10.1007/s00530-024-01617-0

Zheng, J., Zhao, R., Yang, G., Liu, Y., Zhang, Z., Fu, Y., et al. (2023). An underwater image restoration deep learning network combining attention mechanism and brightness adjustment. *J. Mar. Sci. ENG.* 12, 7. doi: 10.3390/jmse12010007

Zhu, X., Lyu, S., Wang, X., and Zhao, Q. (2021). "TPH-YOLOv5: improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*. 2778–2788. doi: 10.48550/arXiv.2108.11539

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., and Dai, J. (2020). Deformable DETR: deformable transformers for end-to-end object detection. *arxiv preprint arxiv:2010.04159*. doi: 10.48550/arXiv.2010.04159