ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0013

Research Review of Underwater Target Detection Technology based on Path Aggregation Network

Peihan Yang, Haifeng Yu*, Zonglin Li, Liangjing Fan, Wenzewei Liu

North China University of Science and Technology, Tangshan, 063000, China

Abstract

Underwater target detection technology is of great significance in marine research and applications, but its performance is severely constrained by the complexity of the underwater environment, such as light attenuation, scattering and degradation of image quality due to turbid water. Traditional methods are limited in their effectiveness due to the difficulty in dealing with multi-scale targets and complex backgrounds. In recent years, deep learning-based path aggregation network (PANet), as an efficient feature fusion structure, is able to effectively fuse multi-scale features and enhance the target detection performance through top-down and bottom-up path design. This paper systematically reviews the technical principles of PANet and its current research status in underwater target detection, then focuses on the analysis of the improved PANet model based on YOLO series, and discusses the challenges faced by this technology in the directions of model lightweighting, data diversity enhancement and multimodal fusion. Future research needs to further optimize the computational efficiency, expand the dataset size and explore the cross-modal data fusion, in order to promote the practical application of underwater target detection technology in complex ocean scenarios.

Keywords

Underwater Target Detection; Path Aggregation Network; YOLO; Attention Mechanism.

1. Introduction

In recent years, with the rapid development of marine scientific research, marine environment monitoring and Marine resources development, the importance of underwater target detection technology as a key supporting technology has become increasingly prominent[1]. However, the underwater environment is complicated, and the light propagates in the water will be severely attenuated, resulting in low contrast and color distortion[2] of the image, while the suspended particles in the water will produce scattering and absorption, further interfering with the target detection. In this complex environment, the performance of traditional target detection methods is greatly reduced, so it is urgent to study the efficient target detection technology suitable for underwater environment.

In order to overcome the above challenges, deep learning technology, especially the object detection algorithm based on convolutional neural network (CNN), has made remarkable progress[3] in the field of underwater object detection in recent years. However, the traditional CNN model usually adopts a single feature extraction path, which is difficult to fully capture the diversified feature information of underwater targets. In addition, the scale and attitude of underwater targets vary greatly, and the feature representation of a single scale is difficult to meet the needs of practical applications.

To solve the above problems, the underwater target detection technology[4] based on path aggregation network (PANet) comes into being. By introducing the multi-scale feature fusion

Volume 11 Issue 5, 2025
DOI: 10.6919/ICJE.202505_11(5).0013

mechanism, PANet can effectively solve the shortcomings of the traditional CNN model in feature extraction and scale adaptability, and bring a new solution for underwater target detection.

2. Path Aggregation Network

2.1 Overview

The Path Aggregation Network (PANet) is based on the feature Pyramid Network[5] (FPN). FPN builds a feature pyramid through top-down and bottom-up paths, fusing feature maps of different levels to improve the model's ability[6] to detect objects at different scales. However, FPN does not make full use of the details in the low-resolution feature maps during the feature fusion process. PANet adds a bottom-up path enhancement, which can better retain the details in the low-resolution feature map through information transfer and fusion between different levels, so that the model can locate the target more accurately in the target detection task. The approximate structure is shown in Figure 1:

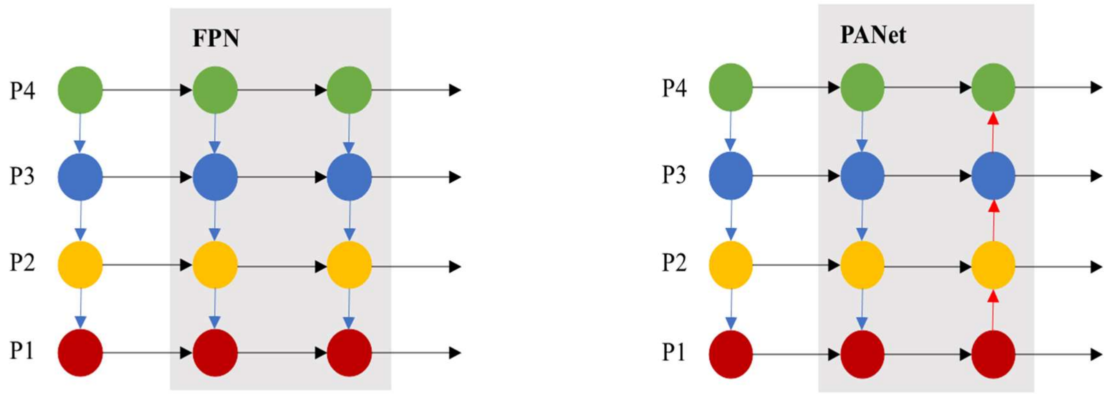

Figure 1. Comparison of PAN and PANet structures

2.2 Research Status

Path aggregation network (PANet) was originally proposed by Liu[4] et al in 2018 for instance segmentation. The core idea is to enhance the expression of feature pyramid by introducing bottom-up and top-down path aggregation mechanisms, so as to improve the accuracy of multi-scale target detection.

Subsequently, its bidirectional feature pyramid structure was quickly introduced into the field of target detection. Early studies mainly focused on combining PANet with traditional detection frameworks (such as Faster R-CNN and YOLO) to improve the accuracy of underwater target detection. For example, in 2020, Liu T[7] et al. introduced the improved cross PANet on the basis of YOLOv3. It is a cross-feature map feature fusion structure, which can effectively integrate semantic and positional information in multi-scale feature maps, thus improving the performance of target detection. Liu J[8] et al proposed an improved underwater target detection algorithm based on Faster R-CNN in 2022: Swin Transformer architecture is used to replace the original backbone network to strengthen feature extraction, and cross-layer feature interaction mechanism is constructed by bidirectional feature pyramid. Experiments show that the improved model significantly improves the recognition robustness of multi-scale targets in complex underwater scenes. With the rapid development of deep learning technology, PANet's improved model has been continuously deepened in the field of underwater target detection. Some researches introduce attention mechanism to cope with underwater background interference. For example, Zhang et al. [9]innovationally combined cross-level feature fusion strategy with self-attention mechanism by integrating multi-head self-attention module of context awareness and multi-scale feature interaction mechanism, and realized efficient underwater environment detection based on YOLOv5s architecture. The other direction is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0013

to combine PANet with underwater image enhancement algorithm; In addition, lightweight design has become a research hotspot in underwater real-time detection. Zhang et al[10.] built a lightweight underwater target detection strategy based on MobileNet v2 to solve the problems of high complexity and poor real-time performance of underwater target detection models, and introduced attention mechanism in FPN network. AFFM module is used to improve the feature fusion mode, better deal with semantic and scale differences, and achieve improved accuracy, so as to achieve the balance between speed and accuracy of target detection. It is worth mentioning that in 2023, Yu et al. [11] proposed a new model, Multi-attention path Aggregation Network (APAN), which combines multiple attention mechanisms and optimizes the structure of the path aggregation network. The multi-attention mechanism can adaptively focus on different regions and feature channels, enhancing the extraction [12]of important target features; Path aggregation network optimizes the propagation path of features at different levels, so that the network can make better use of multi-scale information[13]. The combination of the two methods in underwater target detection is expected to break through the bottleneck of existing technologies and improve detection accuracy and robustness. In addition, Yu et al. also combined PANet with image enhancement technology to form a multi-stage optimization framework, and then input the enhanced images into PANet for detection, which greatly improves the detection and accuracy of the dataset in turbidity waters. At the same time, compared with similar models, APAN ensures accuracy while reducing parameter redundancy through lightweight attention module design. For example, through pruning and quantitative compression, the inference efficiency of the model in embedded devices is improved by about 18%, providing a feasible technical solution for real-time detection of underwater robots.

However, it is difficult for a single optical image to cope with the complex underwater environment, so multi-modal data fusion has become a research hotspot. PANet is widely used in optical-sonar[14] cross-mode detection because of its strong feature fusion capability.

In addition to underwater images, PANet has also been widely used in medical[15] images[16], remote sensing images, automatic driving [17]and other specific scenarios. For example, in the medical image segmentation task, PANet improves the detection accuracy of focal areas through multi-scale feature fusion; In remote sensing image analysis, PANet is used to optimize multi-spectral feature fusion and improve the accuracy of target recognition. The field of automatic driving relies on the improved real-time detection performance of PANet, which is used to identify pedestrians, vehicles and traffic signs in complex scenes. These applications further verify the versatility and effectiveness of PANet in multi-scale feature extraction and fusion.

However, although PANet and its improved model have made remarkable progress in underwater target detection, there are still some problems that need to be solved. For example, the high complexity of the model and the aggregation of dynamic attention mechanisms and dual-path features increase the computational burden, limiting its application in scenarios with high real-time requirements. In addition, the underwater target detection dataset is limited in size and diversity, which is difficult to cover the changes of different sea areas, light conditions and seasons, limiting the generalization ability of the model. Future studies need to further explore lightweight design, data enhancement and multi-modal fusion techniques to improve the performance and adaptability of path aggregation networks in practical applications.

3. Improved Model of PAth-aggregation Network

3.1 Dataset

The performance evaluation of underwater target detection is highly dependent on the diversity and representativeness of the datasets[18]. We divide the datasets into public datasets and custom datasets. These datasets usually contain images or videos taken underwater and mark different types of underwater targets (such as fish, coral, shipwrecks, artificial facilities, etc.). As shown in Table 1 and Figure 2, the following are some common underwater target detection datasets and related legends:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0013

Figure 2. Examples of common underwater target detection datasets

Table 1. Common underwater target detection data sets

Data set Name	Year	Number of categories	Major category goals	Number of images
DUO[19]	2021	4	Sea cucumber, sea urchin, scallop, starfish	7782
URPC2021[20]	2021	4	Sea cucumber, sea urchin, scallop, starfish	8800
Fish4Knowledge[21]	2013	23	23 species of fish	27370
TrashCan[22]	2020	22	Bottom litter, Marine life	7212
SUIM[23]	2020	8	A variety of Marine life	1635
Kyutech10K[24]	2018	7	A variety of Marine life	10728

3.2 An Improved Model based on the YOLO Series

YOLO (You Only Look Once) [25] series models, as the classic models in the field of target detection, have attracted wide attention for their fast and efficient detection capabilities. In order to improve the performance of YOLO model in underwater target detection, researchers introduced PANet to improve it. According to the particularity of underwater environment, they put forward a variety of improvement strategies to optimize the PANet module of YOLO model.

For example, Hu et al[26] proposed an improved scheme based on YOLOv4 to solve the problems of low quality images, small target recognition and model redundancy faced by feed particle detection in aquaculture: reconstructing FPN+PANet feature pyramid connection mode, and using fine-grained feature map to enhance the capturing ability of very small particles. DenseNet was introduced to optimize the residual structure of CSPDarknet, and the feature reuse efficiency was improved through dense connection, while the pruning redundancy parameters were used to balance the accuracy and calculation cost. Experiments show that the improved model greatly improves the detection accuracy in real farming scenarios, increasing the mAP from 65.40% to 92.61%, while reducing the calculation amount by 30%, verifying its efficiency and practicability in high-density and small-target underwater detection tasks, and providing reliable technical support for intelligent feeding system. Aiming at the challenges of fuzzy underwater targets and insufficient feature resolution, Lei et al. [27] introduced Swin Transformer into YOLOv5 as the backbone network, combined with PANet method optimization, optimized the multi-scale fusion effect by focusing on important resolution features, and adjusted the confidence loss function to strengthen the learning ability of high-quality positive

ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0013

anchor frame. Experimental results show that the improved model performs well in complex underwater environments, significantly outperforming traditional target detection algorithms, and providing reliable technical support for ocean exploration missions. To solve the problem of Marine benthos detection, Zhang[28] et al. proposed YOLOXT, a new model, which makes a breakthrough through FPST-PAN structure: multi-scale jump mechanism and Swin Transformer are integrated into the improved path aggregation network to strengthen cross-level feature interaction capabilities, and complex target perception is enhanced with deformable attention. Experiments show that the mA of IOC-URPC dataset improved by 3.9% compared with that of YoloX, which effectively supports the quantitative identification task of Marine organisms, and validates the core value of reconstructed PANet for underwater multi-scale and diverse target detection. Aiming at problems such as image blur and feature loss in underwater robot target detection, Chen et al. [29] proposed an improved algorithm based on YOLOv7. By constructing Cat-BiFPN structure and combining PANet to achieve weighted nonlinear feature fusion, the problem of fine-grained information loss is effectively solved. Experiments show that the accuracy rate and recall rate of the proposed algorithm on UPRC offshore dataset are improved by 2.9%, while the number of model parameters is reduced by 11.2%, which significantly improves the detection performance and efficiency. Aiming at the storage and computing power problems of underwater vehicles, Zhao et al. [30] proposed a lightweight model FEB-YOLOv8. Based on the YOLOv8 framework, the model is improved in many aspects. By combining FPN+PAN strategy in the Neck part and fusing deep and shallow semantic information through cross-layer connection, they developed a new feature pyramid network architecture, successfully achieving an optimal trade-off between model lightweight and detection accuracy. Experiments show that the improved model can improve the accuracy of underwater data set, and the calculation amount and parameters are greatly reduced, which is a favorable scheme for underwater object detection. For Marine flexible biological detection, Tian et al. [31] improved the PAN module of YOLOv8n to RepBi-PAN, optimized the feature fusion path, and combined SimAM attention mechanism to improve the key feature extraction capability. Meanwhile, deformable convolution (DCN) was integrated to enhance deformable feature capture, and WIoU loss function was used to optimize the anchor frame quality. Experiments show that the improved model significantly improves detection accuracy compared with the traditional YOLOv8n, and provides an efficient solution for flexible biomonitoring.

3.3 Summary of Experimental Results

The following table summarizes the relevant contents of the above experiments:

Table 2. Improved YOLO series model

Literature	YOLO Model	Data set	mAP(%)	Application scenarios
[26]	v4-improve	Custom data sets	92.6	Feed particle Monitoring
[27]	v5-improve	URPC	87.2	Underwater Exploration Inspection
[28]	v7-improve	UPRC offshore	66.6	Underwater robot Inspection
[29]	YOLOXT	IOC-URPC	70.9	Benthic biometrics
[30]	v8-improve	DUO	82.9	Underwater light weight detection
		URPC2020	83.5	
[31]	v8n-improve	Custom data sets	80.5	Flexible biomonitoring

4. Shortcomings and Challenges of Current Research

Although PANet has made remarkable progress in underwater target detection, there are still some challenges:

Model complexity and computational efficiency: Although the improved path aggregation structure significantly improves the detection accuracy, it also greatly increases the computational complexity and training difficulty of the model. For example, the dynamic adjustment of the attention module and dual path feature aggregation in APAN require additional computational resources, which may result in significantly longer training and inference time for the model. In addition, highly complex network structures are limited in underwater application scenarios that require high real-time performance, such as underwater robot navigation. Therefore, how to optimize the network structure and reduce the calculation cost without reducing the detection accuracy is one of the urgent problems to be solved at present.

Data dependency and dataset limitations: Deep learning models usually rely on a large number of labeled data for training, but underwater target detection datasets are relatively scarce, and the cost of labeling is high[32]. Although the existing underwater image data covers some common Marine objects, its diversity and representation are still insufficient, and it is difficult to cover the complexity of the actual underwater environment. For example, underwater images can vary significantly across ocean areas, light conditions and seasons, and existing datasets may not fully reflect these variations. In addition, the limited size of the datasets also limits the model's ability[33] to generalize. Therefore, how to expand the size of data set, improve the diversity of data, and reduce the dependence of model on data is an important direction of current research.

The complexity of underwater environment: the poor quality of underwater image is an important factor affecting the accuracy of target detection. The problems of light refraction, scattering and water turbidity cause the underwater image blur, low contrast and color distortion. In addition, the underwater target is often small and the background is complex, the contrast between the target and the background is low, and it is easy to be blocked. Although some studies have introduced image enhancement techniques and multi-attention mechanisms to improve feature extraction capabilities, they still face the problem of insufficient detection accuracy under extreme conditions (such as low light and cloudy waters in deep sea). Therefore, how to further optimize the image enhancement technology and improve the robustness of the model to the complex underwater environment is one of the key challenges of current research.

Lack of model generalization: Although the improved PANet performs well on specific data, its adaptability to new environments and unknown targets is poor. For example, the detection performance of the model may be significantly reduced when faced with light conditions in different seas or new Marine biological targets. In addition, the model's lack of robustness under extreme conditions also limits its application in real underwater scenarios. Therefore, how to improve the generalization ability[34] of the model so that it can better adapt to the diversified underwater environment is a problem that current research needs to focus on.

Lack of multi-modal data fusion: The current PANet model mainly relies on optical images for target detection, but optical images have many limitations in underwater environment. For example, in cloudy waters or under low light conditions, the quality of optical images may be seriously degraded, affecting the detection accuracy[35]. In contrast, multi-modal data such as sonar and infrared have unique advantages in underwater environments, but the application[36] of multi-modal data fusion in underwater target detection has not been fully explored in current research. Therefore, how to realize the effective fusion of multi-modal data and solve the problem[37] of heterogeneity between different modal data is an important direction to improve the underwater target detection performance.

5. Summary and Prospect

This paper systematically reviews the research progress of underwater target detection technology based on path aggregation network (PANet). Aiming at the challenges of complex underwater environment, poor image quality and diverse target scales, PANet integrates multi-scale features through bidirectional feature pyramid structure, which significantly improves the detection capability of small targets and robustness under complex background. The improved model further optimizes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0013

detection performance and efficiency by introducing strategies such as dynamic feature fusion, multi-attention mechanism, and lightweight design. Experiments show that these models achieve significant accuracy improvement on datasets such as URPC and IOC-URPC, and show application potential in practical scenarios such as feed particle monitoring, benthic biometrics, and underwater robots. Although progress has been made in the current research, challenges remain. In the future, the application of technologies such as lightweight design, data enhancement and multi-modal fusion is expected to further improve the performance of underwater target detection technology, expand its application in Marine scientific research, underwater archaeology and resource development, and provide more powerful technical support for Marine industry.

Acknowledgments

This work is supported by Research Project of North China University of Science and Technology College Students Innovation and Entrepreneurship Training Program Project, Project Number: X2024083.

References

[1] Lu Jiahao. Research on super-resolution algorithm of underwater image based on improved channel attention residual network [D]. Guilin University of Electronic Science and Technology,2023.

[2] Qian Junbing, Jiang Liangwei, Quan Haiyan, et al. An underwater image enhancement method combining multi-channel equalization and multi-scale fusion [J]. Optoelectronics. Laser, 2022, 33(4):10.

[3] Dhillon A, Verma G K. Convolutional neural network: a review of models, methodologies and applications to object detection[J]. Progress in Artificial Intelligence, 2020, 9(2): 85-112.

[4] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[5] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[6] Chen Mengxing, Chen Wei, Chen Guojun, et al. Lightweight Algorithm of Underwater human sonar image real-time target Detection based on YOLOv8 [J]. Light Industry Science and Technology, 2024, 40 (06): 111-114.

[7] Liu T, Pang B, Ai S, et al. Study on visual detection algorithm of sea surface targets based on improved YOLOv3[J]. Sensors, 2020, 20(24): 7263.

[8] Liu J, Liu S, Xu S, et al. Two-stage underwater object detection network using swin transformer[J]. IEEE Access, 2022, 10: 117235-117247.

[9] Zhang Y, Liu T, Yu X. Contextual and lightweight network for underwater object detection with self-attention mechanism[C]//2023 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2023: 1644-1649.

[10] Zhang M, Xu S, Song W, et al. Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion[J]. Remote Sensing, 2021, 13(22): 4706.

[11] Yu H, Li X, Feng Y, et al. Multiple attentional path aggregation network for marine object detection[J]. Applied intelligence, 2023, 53(2): 2434-2451.

[12] Zou Ziyin, Gai Shaoyan, Da Feipeng, et al. Acta Optica Sinica, 2021, 41(15): 1515001.] Acta Optica Sinica, 2013, 41(15): 1515001.

[13] Chi Y, Li J, Fan H. Pyramid-attention based multi-scale feature fusion network for multispectral pan-sharpening[J]. Applied Intelligence, 2022, 52(5): 5353-5365.

[14] Li Z, Xie Z, Duan P, et al. Dual spatial attention network for underwater object detection with sonar imagery[J]. IEEE Sensors Journal, 2024, 24(5): 6998-7008.

[15] He D, Li Y, Chen L, et al. A two-stage image enhancement and dynamic feature aggregation framework for gastroscopy image segmentation[J]. Neurocomputing, 2024, 601: 128228.

[16] Dang L, Huangfu P, Hou Y, et al. A path aggregation network based on residual feature enhancement for object detection in remote sensing imagery[J]. Remote Sensing Letters, 2023, 14(6): 598-608.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0013

[17] Fang S, Zhang B, Hu J. Improved mask R-CNN multi-target detection and segmentation for autonomous driving in complex scenes[J]. Sensors, , 2023, 23(8): 3853.

[18] Fu C, Liu R, Fan X, et al. Rethinking general underwater object detection: Datasets, challenges, and solutions[J]. Neurocomputing, 2023, 517: 243-256.

[19] Liu C, Li H, Wang S, et al. A dataset and benchmark of underwater object detection for robot picking[C]//2021 IEEE international conference on multimedia & expo workshops (ICMEW). IEEE, 2021: 1-6.

[20] Cai S, Li G, Shan Y. Underwater object detection using collaborative weakly supervision[J]. Computers and Electrical Engineering, 2022, 102: 108159.

[21] Fish4Knowledge: collecting and analyzing massive coral reef fish video data[M]. Berlin/Heidelberg, Germany: Springer, 2016.

[22] Hong J, Fulton M, Sattar J. Trashcan: A semantically-segmented dataset towards visual detection of marine debris[J]. arXiv preprint arXiv:2007.08097, 2020.

[23] Islam M J, Edge C, Xiao Y, et al. Semantic segmentation of underwater imagery: Dataset and benchmark[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 1769-1776.

[24] Lu H, Uemura T, Wang D, et al. Deep-sea organisms tracking using dehazing and deep learning[J]. Mobile Networks and Applications, 2020, 25: 1008-1015.

[25] Jiang P, Ergu D, Liu F, et al. A Review of Yolo algorithm developments[J]. Procedia computer science, 2022, 199: 1066-1073.

[26] Hu X, Liu Y, Zhao Z, et al. Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network[J]. Computers and electronics in agriculture, 2021, 185: 106135.

[27] Lei F, Tang F, Li S. Underwater target detection algorithm based on improved YOLOv5[J]. Journal of Marine Science and Engineering, 2022, 10(3): 310.

[28] Zhang J, Yongpan W, Xianchong X, et al. YoloXT: A object detection algorithm for marine benthos[J]. Ecological Informatics, 2022, 72: 101923.

[29] Chen L, Yin T, Zhou S, et al. Underwater target detection algorithm based on feature fusion enhancement[J]. Electronics, 2023, 12(13): 2756.

[30] Zhao Y, Sun F, Wu X. FEB-YOLOv8: A multi-scale lightweight detection model for underwater object detection[J]. PLoS One, 2024, 19(9): e0311173.

[31] Tian Y, Liu Y, Lin B, et al. Research on marine flexible biological target detection based on improved YOLOv8 algorithm[J]. PeerJ Computer Science, 2024, 10: e2271.

[32] Meng Xiao-Xuan. Research on object detection Algorithm based on deep Active Weakly Supervised Learning [J]. Information and Computer (Theoretical Edition), 2021, 33 (02): 61-64.]

[33] Zech J R, Badgeley M A, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study[J]. PLoS medicine, 2018, 15(11): e1002683.

[34] Zhang C, Bengio S, Hardt M, et al. Understanding deep learning (still) requires rethinking generalization[J]. Communications of the ACM, 2021, 64(3): 107-115.

[35] Li Ming-gui, Zhou Huan-yin, Gong Li-wen. Application of ROV underwater obstacle detection and obstacle avoidance technology [J]. Computer Engineering and Applications, 2024, 60 (17): 34-47.

[36] Chen H, Wang Z, Qin H, et al. UAMFDet: Acoustic-Optical Fusion for Underwater Multi-Modal Object Detection[J]. Journal of Field Robotics, 2024.

[37] Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects[J]. Proceedings of the IEEE, 2015, 103(9): 1449-1477.