ARTICLE
https://doi.org/10.26599/AIR.2026.915010
CAAI Artificial Intelligence Research

Underwater Small Object Detection: A Survey

Peixian Zhuang[1✉,*], Yuchun Wang[1,*], Yihang Wang[1], Fei Liu[1], Yanchen Guo[2], and Zhenqi Fu[3]

ABSTRACT

Underwater small object detection is one of the emerging core technologies at the intersection of computer vision and underwater imaging detection, which aims to achieve efficient and accurate detection and identification of faint, small targets in underwater environments, and has been widely applied in fields of aquatic organism detection, mineral resource exploration and other fields. Among them, underwater object detection methods based on optical images, with their high resolution and good flexibility, have demonstrated significant advantages in underwater biological detection, recognition and other scenarios. However, there is currently a relative lack of comprehensive reviews on existing research regarding underwater small object detection. To address this gap, this paper presents a systematic review of small object detection in underwater optical images. Firstly, at the level of problems and data, the specific problems faced by underwater small detection are comprehensively sorted out and classified into two categories: the inherent characteristics of underwater small objects, and the degradation characteristics of underwater images. Then 16 existing typical datasets are sorted out and analyzed in detail. Subsequently, at the methodological level, existing underwater small object detection algorithms are systematically summarized from the perspectives of traditional machine learning and deep learning, and an in-depth analysis of improved deep learning-based strategies is conducted by covering five aspects: data augmentation and pre-processing, multi-scale feature fusion, attention mechanism introduction, loss function optimization, and lightweight design. Finally, the challenges faced by underwater small object detection at present are pointed out, and the potential directions for future development are explored, aiming to provide valuable references for subsequent related research.

KEYWORDS

Underwater scenarios, Optical images, Small object detection, Datasets, Deep learning.

1 Introduction

The ocean gives birth to life, connects the world, and promotes development. The ocean harbors abundant biological and mineral resources, whose development and utilization hold irreplaceable strategic significance for promoting sustainable national economic development. Caring for the ocean, understanding the ocean, managing the ocean, and building a maritime power are major strategic tasks to realize the great rejuvenation of the Chinese nation. With the growing demand for intelligent underwater imaging, underwater object detection technology has been extensively studied in numerous scenarios, including aquatic organism detection, mineral resource exploration, marine salvage operations, etc. In underwater environments, optical methods demonstrate significant advantages for close-range underwater tasks such as biological identification, as their high resolution provides rich details for target recognition and behavior analysis[1, 2], while offering considerable operational flexibility. However, underwater small object detection remains particularly challenging, and its performance has consistently lagged behind that for larger and medium-sized objects, constrained by objects' small size, limited extractable features, and complexities of underwater environments. In consequence, boosting the performance of underwater small object detection has progressively emerged as a key research focus.

In recent years, various algorithms of underwater small object detection have developed from traditional machine learning methods to deep learning-based approaches, and some reviews are summarized in this field. Rout *et al.*[3] reviewed current underwater monitoring from three aspects: enhancement, object detection, and object tracking, but paid limited attention to underwater small objects. Wei *et al.*[4] discussed challenges and solutions for small object detection in applications of automatic driving, vehicle-based object detection, and aerial image detection, but not focused on underwater scenes. Xu *et al.*[5] provided a systematic analysis of deep learning-based underwater object detection, along with discussing its current challenges, future trends, and potential applications, and further explored intrinsic correlation between underwater

1 Mariner Lab, Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
2 Laoshan Laboratory, Qingdao 266237, China
3 Department of Automation, Tsinghua University, Beijing 100084, China
Address correspondence to Peixian Zhuang, zhuangpeixian0624@163.com
∗ These authors contributed equally to this work.

image enhancement and underwater object detection. Dong *et al.*[6] presented underwater object detection datasets, and exhibited that directly applying image enhancement did not effectively improve the detection accuracy, whereas high-resolution networks could enhance the performance. Based on these existing achievements, Table 1 summarizes related research of underwater small object detection from four perspectives, where '√' indicates that the article provides a detailed summary and analysis in that aspect. 'Background' indicates whether an analysis of small objects is based on underwater or non-underwater scenarios. 'Problem' represents whether the issue faced in underwater small object detection is analyzed and summarized. 'Methodology' can be divided into two classes: 'Traditional Methods' and 'Deep Learning Methods', where traditional methods refer to machine learning-based methods based on handcrafted features. 'Dataset' indicates whether existing datasets of underwater small object detection are summarized and analyzed.

Table 1 Existing surveys on underwater small object detection

Year	Ref.	Background		Problem	Methodology		Dataset
		Non-underwater	Underwater		Traditional	Deep learning	
2025	[7]	✓		✓		✓	
2025	[8]		✓			✓	✓
2025	[9]		✓		✓	✓	✓
2024	[3]		✓			✓	✓
2024	[4]	✓		✓		✓	✓
2023	[10]		✓		✓	✓	
2023	[5]		✓		✓	✓	
2023	[11]	✓		✓	✓	✓	
2022	[6]		✓		✓	✓	✓
2022	[12]		✓		✓	✓	
2020	[13]		✓		✓	✓	
2017	[14]		✓		✓	✓	

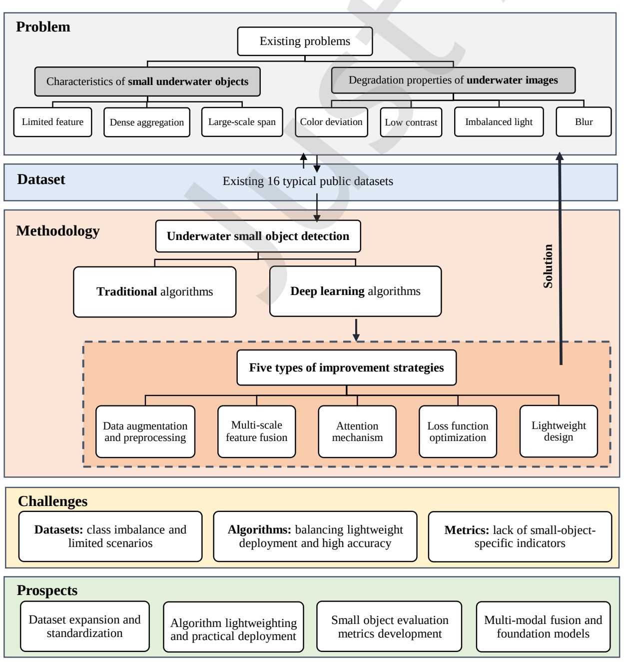

Fig. 1 Organizational framework of the article.

Note that the above existing reviews mainly focus on underwater large-sized and medium-sized objects detection or small objects in natural scenes, generally lacking a systematic overview of algorithms specifically designed for underwater small object detection. Besides, the summaries of public datasets for underwater small object detection are typically inadequate in coverage. To this end, we propose a systematically comprehensive survey of underwater small object detection. The organizational framework of our article is overviewed in Fig. 1. Starting from the problem posed by the dual constraints of small object size and underwater environment in Sec. 2, existing public datasets and detection methods based on traditional machine learning and deep learning are systematically summarized and analyzed in Sec. 3 and Sec. 4 respectively. In Sec. 5, current challenges about datasets, algorithms and metrics are pointed out, and potential directions for future development are explored in this field.

The main contributions of this article are summarized as follows: (1) To the best of our knowledge, this is the first work on reviewing underwater small object detection, and a comprehensive summary is presented from five dimensions: problems, datasets, methodology, challenges and prospects. (2) We condense out the core issues arising from inherent characteristics of underwater small object and underwater image degradations. (3) We conduct a multi-dimensional analysis of existing sixteen typical datasets of underwater small object detection, meanwhile, summarize existing detection algorithms based on traditional machine learning and deep learning, with placing particular emphasis on discussing five categories of improvement strategies for deep learning. (4) We suggest current challenges associated with datasets, algorithms and metrics, and further point out insightful directions for future research.

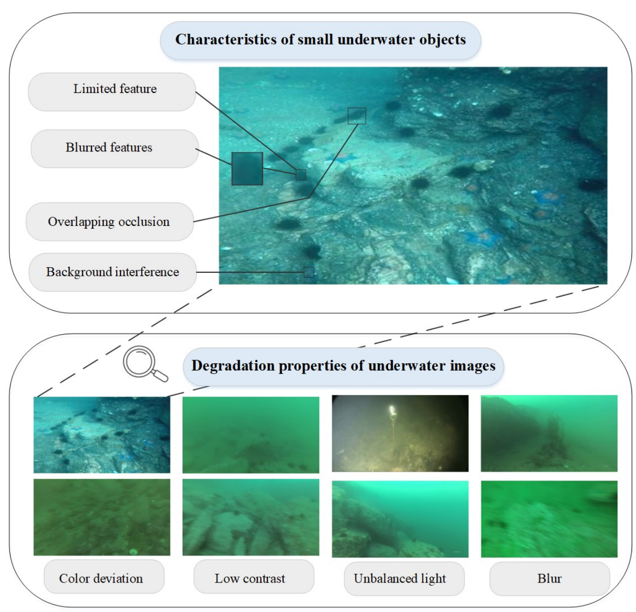

Fig. 2 Core problems of underwater small object detection. The example images are sourced from the three datasets Brackish [25], DUO [28] and UODD [29].

2 Problem

The complexity of underwater environment, compounded by inherently small sizes of objects, poses numerous challenges for subsequent underwater detection tasks. As shown in Fig. 2, this section categorizes the core issues of underwater small object detection from the two dimensions: characteristics of underwater small objects, and degradation properties of underwater images.

2.1 Characteristics of small underwater objects

In the task of object detection, there is no unified standard for defining a small object. The current definitions are established based on two aspects: absolute scale and relative scale. Among them, the absolute scale is determined by whether the object's pixel dimensions fall below a certain threshold, which varies across different datasets. For instance, in the widely used MS COCO dataset [15], the objects with a resolution smaller than 32 pixels $\times$ 32 pixels are defined as small objects. The relative scale is based on the object's proportion within an image, typically where the object's pixel area constitutes less than 1% of the total image. Another study [16] defines the objects with an area-to-image ratio between 0.08% and 0.58% as small objects. Since underwater image resolutions vary substantially across different datasets, the definition using fixed pixel-based criterion is not directly suitable for cross-dataset comparison. To address the above issue, Tsai *et al.* [17] adopted relative object area as the scale criterion for underwater dataset DUO, and classified objects occupying 0–0.25% of the image area as small objects. Following the above relative-scale criterion, this survey adopts the threshold 0.25% for calculating the proportion of small objects across different underwater datasets. In underwater small object detection, these targets usually exhibit the following characteristics.

(a) Small size with limited features (b) Dense aggregation (c) Large-scale span

Fig. 3 Characteristics of underwater small objects. The example images are sourced from the three datasets Brackish [25], DUO [28] and LSUI [31].

Limited feature. Small objects occupy a low pixel proportion in images, and possess limited features such as image contours and textures [18]. In underwater environments, as labeled in Fig. 3(a), small marine organisms like sea urchins or fish fry consist of only a few dozen or a handful of pixels, which results in extremely scarce visual features. It is challenging to preserve effective features of small objects after multiple pooling operations in Convolutional Neural Networks (CNNs), since their information diminishes in feature maps [19]. This issue is further exacerbated in underwater settings, where factors such as water current disturbances and light refraction often cause underwater small objects to lack distinguishable structures and textural details.

Dense aggregation. Underwater small objects generally appear in dense aggregation. As shown in Fig. 3(b), examples involve schools of juvenile fish. The inherently small size of these objects, combined with the occlusion from overlapping individuals, presents two main challenges. First, it is difficult for existing algorithms to distinguish individual boundaries, and anchor boxes suffer from cross-matching, leading to duplicate detections or missed objects. Second, category confusion arises, when small objects of the same category are densely packed, underwater detection algorithms tend to mistakenly identify adjacent objects as a single entity. Furthermore, when different categories of small objects aggregate—such as shrimp and small fish swimming together—misclassification is likely to occur.

Large-scale span. The pixel scale of underwater objects spans a wide range, from small fish measuring a dozen pixels to large fish occupying several tens of pixels. As illustrated in Fig. 3(c), when small and large objects coexist, deep learning-based detection models struggle to adapt their receptive fields [20]. For small objects, an overly large receptive field leads to be overwhelmed by the background information. For the objects of medium scale, an insufficient receptive field fails to capture complete features, resulting in inconsistent detection performance across different scales. For instance, in an image containing both a sea turtle and small fish, when the small fish is near the turtle, detection models easily misclassify it as the background noise. Moreover, the insufficient receptive field for the sea turtle hinders adequate feature extraction. Consequently, it is challenging to achieve precise detection performance across all scales under a unified model architecture.

2.2 Degradation properties of underwater images

The quality of underwater images is closely linked to the level of detection performance. High-quality underwater images, characterized by clear details, good contrast, and minimal noise interference, aid in object distinguishability, feature extraction, and algorithm generalization. However, in underwater small object detection, degradation properties of underwater images are shown below.

(a) Color deviation (b) Low contrast (c) Imbalanced light (d) Blur

Fig. 4 Degradation properties of underwater images. The example images are sourced from the three datasets Brackish [25], DUO [28] and UODD [29].

Color deviation. The core reason for the degradation of underwater image quality lies in complex physical processes of light propagation in water, with interference from the water medium being another critical factor. Specifically, the water exhibits selective absorption of light at different wavelengths. The red light, attenuates most rapidly due to its longer wavelength, while green and blue lights attenuate more slowly [21] with their shorter wavelengths. This results in color deviations of underwater images, such as shown in Fig. 4(a) where underwater images exhibit blue or green tone.

Low contrast. The background light generated by the back-scattering elevates overall brightness of underwater images, whereas the light reflected from objects is attenuated and dimmer. The combined effect of these factors compresses the grayscale difference between the object and the background, leading to a significant decrease in image contrast, as illustrated in Fig. 4(b). The forward-scattering causes the object's light to diffuse, resulting in blurring edges and losing details.

Unbalanced light. The intensity of natural light decays exponentially with increasing water depth, causing insufficient image brightness [22]. As exhibited in Fig. 4(c), artificial lights create a brightness gradient that is brighter nearby and darker farther away. Coupled with real-time variations caused by the surface waves, this severely compresses underwater image's dynamic range, leading to the coexistence of overexposed and underexposed regions within the same image [23].

Blur. The motion blur caused by water currents and underwater microorganisms, along with noise from equipment shake—as shown in Fig. 4(d) further exacerbates the deterioration of underwater images. These above problems compound each other, presenting severe challenges for underwater small object detection.

3 Dataset

Underwater object detection datasets serve as a fundamental pillar for providing rich learning samples to support correspondence detection algorithms from training process to practical deployment, and enabling to validate the effectiveness of new methods. Since underwater image datasets of small targets remain scarce, and existing datasets used in underwater object detection include relevant samples of such targets, we conduct a summary and analysis of 16 publicly available underwater image datasets, shown in Table 2, along with their sources and download links.

URPC datasets: Underwater Robot Professional Contest (URPC) has been updated annually since its establishment in 2017. It provides underwater target detection datasets for competition purposes, and includes four main categories: scallop, sea urchin, holothurian, and starfish. The images depict diverse scenes near the site of Zhangzidao and Dalian, and exhibit typical degradations such as color cast and low contrast, making it suitable for robustness evaluation. Across the versions, the four target categories remain unchanged, while the datasets have been refined via changes in image scale, annotation quality, and scene diversity. Later versions include more varied underwater scenes and refined annotations, providing broader conditions for evaluating underwater object detection. For instance, Fig. 5 shows six example images from these datasets.

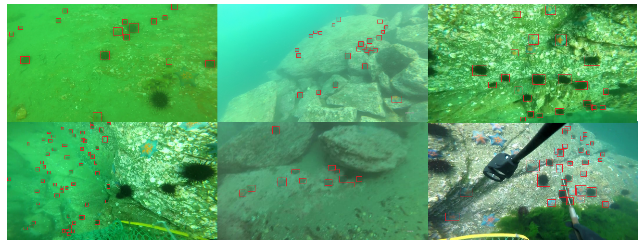

Fig. 5 Examples images from the URPC datasets.

LFW dataset [24]: Labeled Fishes in the Wild (LFW) is sourced from the National Oceanic and Atmospheric Administration (NOAA), consisting of a training-validation set and a test set. The training and validation set is individually divided into two categories: positive samples and negative samples. The positive sample set contains 929 images, focusing on rockfish and associated species in the nearshore rocky seabed environment of Southern California, with a total of 1005 fish targets annotated for location and bounding boxes. The negative sample set includes 3167 fish-free seabed images, used for training algorithms to distinguish 'non-fish' targets.

Four underwater object datasets **UDD** [27], **DUO** [28], **UODD** [29], and **RUOD** [30] have evolved and developed from the URPC dataset. UDD dataset involves 2227 images and provides bounding boxes for three categories: sea urchin, holothurian, and scallop, exhibiting severe imbalance with sea urchins exceeding 90% of instances. DUO dataset is generated by re-annotating the URPC datasets, including 7782 images and consolidating four categories: scallop, starfish, holothurian, and sea urchin, along with sea urchin as the main class. UODD dataset includes 3194 images and is divided into three categories: sea cucumber, sea urchin, and scallop. RUOD dataset is sourced from the URPC datasets and comprises 14,000 images with 74,903 annotated instances across 10 underwater object categories. The above four dataset covers diverse underwater scenarios ranging from shallow sea to coral reefs, and provides specialized test sets for addressing the issues of blur, color cast, and uneven illumination. Fig. 6 exhibits some examples from the four datasets, and their data distributions are shown in Fig. 7.

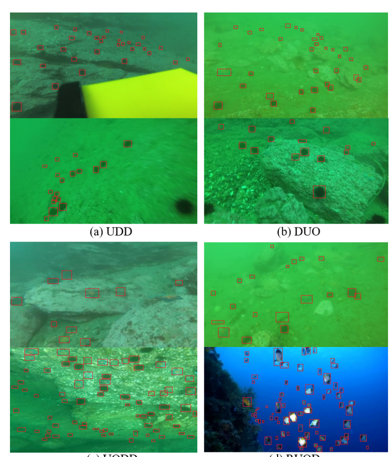

Fig. 6 Examples images from the UDD [27], DUO [28], UODD [29], and RUOD [30] datasets.

Table 2 Summary and quantitative comparison of 16 representative public underwater vision datasets.

Dataset	Year	Ref.	Classes	Images	Annotation	Small object ratio (%)	Avg. relative object size (%)	Object density	Link
URPC 2017	2017		4	18,640	Bounding box	2.64	2.11	18.20	
URPC 2018	2018		4	3,701	Bounding box	12.09	1.20	7.81	
URPC 2019	2019	–	4	5,786	Bounding box	14.89	1.12	7.69	Link
URPC 2020	2020		4	8,975	Bounding box	17.93	1.00	7.04	
URPC 2021	2021		4	8,678	Bounding box	20.78	0.95	7.14	
URPC 2022	2022		4	9,000	Bounding box	23.79	0.91	7.26	
LFW	2015	[24]	–	4,096	Bounding box	8.76	9.71	1.73	Link
Brackish	2019	[25]	6	14,674	Bounding box	36.33	0.72	2.86	Link
SUIM	2020	[26]	8	1,635	Semantic mask	-	-	-	Link
UDD	2021	[27]	3	2,227	Bounding box	45.96	0.66	6.75	Link
DUO	2021	[28]	4	7,782	Bounding box	27.12	0.89	9.58	Link
UODD	2021	[29]	3	3,194	Bounding box	2.72	1.85	6.03	Link
RUOD	2022	[30]	10	14,000	Bounding box	11.39	6.92	5.35	Link
SMD	2024	[31]	40	8,610	Bounding box	26.46	4.27	3.66	Link
COU	2025	[32]	24	9,757	Bounding box + Instance mask	10.49	7.11	1.65	Link
DebrisVision	2025	[33]	24	25,000	Multi-modal annotations	-	-	-	Link

Note: Small objects are defined as instances whose bounding-box area occupies less than 0.25% of corresponding image area, following the relative-scale criterion described in Section 2.1. Average (Avg.) relative object size is calculated as the mean ratio of object area to image area over all annotated instances, while object density denotes the average number of annotated instances per image. For datasets originally providing pixel-level or instance masks, bounding boxes derived from the annotated object regions are used for calculating the scale-related statistics, while the original annotation types are reported in the table. The quantitative statistics are calculated based on the valid annotations in the dataset versions.

Brackish dataset [25]: In Fig. 8 with targeting brackish water environments, this dataset contains 14,674 images from 89 videos with bounding box annotations, and is classified into six categories: big fish, small fish, crab, shrimp, jellyfish, and starfish, collected in Danish straits and featuring high turbidity and complex lighting distinct from open sea. The underwater camera used a 1/3-inch Sony ExView Super HAD Color CCD imaging sensor and a 2.8-mm lens, with LED illumination provided during data acquisition.

SUIM dataset [26]: Semantic Segmentation of Underwater IMagery (SUIM) consists of 1,635 underwater images, with 1,525 images designated for training and validation and 110 images reserved for testing. It covers a variety of scenarios, such as marine exploration and human-robot collaboration, and includes images of multiple resolutions. Eight core semantic categories—fish, reef, aquatic plants, underwater ruins, diver, robot, sea-floor, and background—are defined using a 3-bit binary RGB color encoding scheme. Additionally, this dataset provides annotation masks in two formats: consolidated RGB masks and separate binary masks for each category.

SMD dataset [31]: Seaclear Marine Debris (SMD) dataset consists of 8,610 images, and this data were collected in shallow-water environments using BlueROV2 and SST Mini-Tortuga ROVs equipped with different cameras. Multiple acquisition sites and trials covered different lighting and turbidity conditions. This dataset includes 40 object categories, encompassing not only various types of marine debris such as plastic waste and metal fragments, but also common marine animals and plants. This supports collaborative tasks in underwater litter detection and environmental perception.

COU dataset [32]: Common Objects Underwater (COU) dataset covers two primary environment types: closed water (pool) and open water (lake, ocean), realistically capturing the visual characteristics of different underwater settings, such as variations in lighting and water clarity. All images were collected during underwater robot field trials using a GoPro camera at a resolution of 1920×1080 pixels and a rate of 30 frames per second. The category design focuses on "human-made objects," moving beyond the traditional emphasis on marine life commonly found in underwater datasets. This dataset includes 24 different classes of underwater objects, spanning marine debris, diving tools, and underwater equipment. The data distribution of this dataset is illustrated in Fig. 9.

Debris Vision dataset [33]: This dataset breaks through traditional limitations in scale and composition, comprising a total of 25,000 images divided into two core components. The first component includes 9,430 real underwater images captured from actual marine environments, reflecting real-world conditions such as turbidity variations, lighting fluctuations, and target occlusion. The second component consists of 15,570 synthetic images generated via diffusion models and refined via image translation, covering common underwater debris types as well as less common categories.

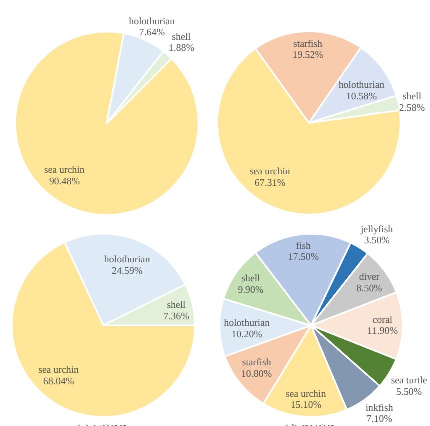

Fig. 7 Data distributions of the UDD [27], DUO [28], UODD [29], and RUOD [30] datasets.

Fig. 8 Examples images from the brackish dataset [25].

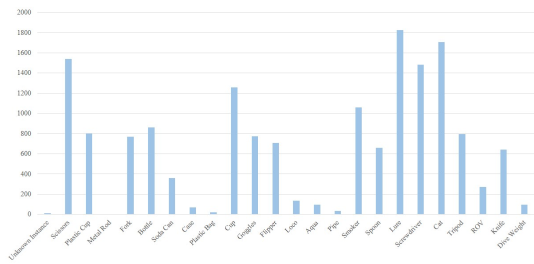

Fig. 9 Data distribution of the COU dataset [32].

Considering quantitative statistics in Table 2 and dataset characteristics, different datasets provide complementary application and evaluation scenarios for underwater vision research. DUO is particularly suitable for underwater detection studies involving small objects and relatively dense object distributions. Brackish is fit for underwater object detection under varying visibility conditions. COU covers different environments, including pools, lakes, oceans, as well as various human-made objects, making it suitable for cross-environment detection of human-made objects. DebrisVision focuses on underwater debris, providing a data basis for underwater debris detection and visual research in complex environments. The above-stated 16 representative public underwater vision datasets provide comprehensive coverage across object categories, including typical marine organisms such as sea urchins, starfish, and holothurian; various types of anthropogenic debris such as plastic bottles and cans; as well as human-operated equipment like diving gear and underwater robotic components, which offers a complete representation of both natural and artificial targets. Moreover, these datasets encompass a wide range of underwater environments, from open waters such as oceans and lakes to enclosed settings like pools. Specific collections, such as Brackish, focus on particular conditions including brackish waters with varying visibility levels. They also involve typical underwater imaging challenges, including blur, low contrast, and similar issues. Additionally, several datasets contain cases of partial occlusion, densely distributed small objects, and multi-category coexisting scenes. In summary, these datasets offer practical data resources for detection tasks and serve as valuable benchmarks to assess algorithm generalization and robustness across different underwater environments.

4 Methodology

This section details existing algorithms of underwater small object detection from traditional and deep learning perspectives, analyzes improvement strategies specifically for underwater small object detection, and summarizes current limitations of various algorithms.

Traditional methodology

Traditional object detection methods can be classified into two main categories: the first class is manually designed feature extractors, such as Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), etc.; the second class is classifier design that employed methods such as Support Vector Machines (SVM), AdaBoost, etc. However, specialized research targeting small objects in underwater images is relatively limited within this methodology. Despite the lack of specialized design for small targets, researchers explored the algorithms of underwater object detection. For detecting slender targets like underwater cables, Fatan *et al.* [34] extracted edge features based on texture information, compared Multi-Layer Perceptron (MLP) and SVM classifiers to suppress background interference, and produced better recognition rate for cable detection tasks. For coral reef fish detection and recognition in underwater HD videos, Villon *et al.* [35] concentrated on supervised machine learning, compared traditional HOG+SVM method with CNN methods on real data, and discuss their strengths and weaknesses. Additionally, for removing uneven illumination and background texture interference in macroalgae recognition, Tan *et al.* [36] introduced an HOG-based scheme combined with an SVM classifier for object detection. For fish detection in underwater sequences, Mathias *et al.* [37] presented a method of underwater object detection, based on Bidimensional Empirical Mode Decomposition (BEMD) and Gaussian Mixture Model (GMM), for probabilistic background modeling, which achieves effective detection of fish targets in complex underwater scenarios. The above traditional methods utilize handcrafted features based on prior knowledge, but are difficult to generalize against complex underwater degradations of scattering, low contrast, and blur. Moreover, detection mechanisms like the sliding window are inefficient. This suggests feature representation deficiency of traditional methodology, and catalyzes a transition toward deep learning methodology.

Deep learning methodology

Deep learning-based approaches have become prevalent for underwater small object detection, and existing deep learning-based underwater object detectors can be broadly grouped into two-stage, one-stage, and Transformer-based frameworks. Two-stage detectors first generated region proposals and then performed object classification and localization, whereas one-stage detectors directly predicted object categories and locations within an unified detection pipeline. Transformer-based detectors used Transformer architectures for feature interaction and adopted query-based mechanisms for object prediction. The further improvements for these methods can be mainly classified into five aspects: data augmentation and preprocessing, multi-scale feature fusion, attention mechanism

Table 3 Comparison of improvement strategies for deep learning algorithms on underwater small object detection

Strategy	Core objective	Typical techniques	Advantages	Disadvantages
Data augmentation and preprocessing	Improve input quality, expand data	Physical model synthesis, GANs, image enhancement	Enhance model generalization, customizable for underwater characteristics	Distribution bias and detail loss
Multi-scale feature fusion	Preserve details of small objects	FPN, PANet, BiFPN, ASFF	Address information loss of small objects, combine details and semantics	Increase computation, fuse background noise
Attention mechanism introduction	targets focus, background suppression	SE, CBAM, Coord Attention, Self-attention	Enhances features of key regions, flexible integration of feature representation	Increases computation, overfit on small datasets
Loss function optimization	Address class imbalance, improve localization accuracy	Focal Loss, IoU-based losses	Tackles training challenges, plug-and-play without architectural changes	Require hyperparameter tuning, depend on high-quality features
Lightweight design	Reduce computation for efficient deployment	Lightweight backbones, pruning, distillation, quantization	Real-time requirements for practical deployment, diverse and mature technical solutions	Accuracy degradation, complex design process

Note: GANs (Generative Adversarial Networks), PANet (Path Aggregation Network), BiFPN (Bidirectional Feature Pyramid Network), ASFF (Adaptively Spatial Feature Fusion).

introduction, loss function optimization, and lightweight design. Table 3 presents their core objectives, typical techniques, advantages, and disadvantages.

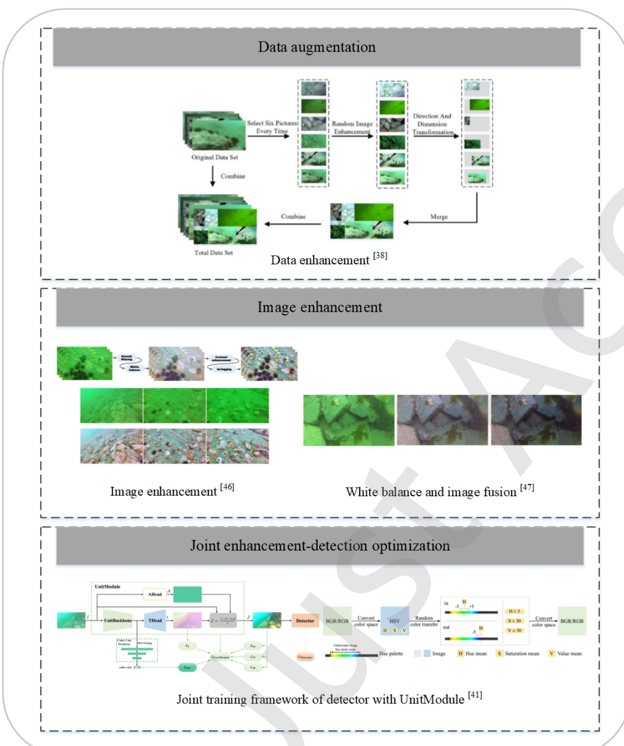

Fig. 10 Examples of data augmentation and preprocessing.

Data augmentation and preprocessing. To address the issues of sample scarcity and image degradation in underwater small object detection, researchers have developed a series techniques of data augmentation and image preprocessing. As illustrated in Fig. 10, these existing methods can be categorized into three main pathways: data augmentation, image enhancement, and joint enhancement-detection optimization. Data augmentation strategies aim to expand and diversify training data, for improving model generalization. Luo *et al.* [38] used a YOLOv4 deep learning network for underwater small object detection, and combined traditional data augmentation methods with the Mosaic algorithm, increasing frequency and diversity of objects and expanding dataset scale. Image enhancement techniques focus on improving visual quality of underwater images. To fast detect underwater fish in underwater videos, Li *et al.* [39] employed a preprocessing algorithm based on Contrast Limited Adaptive Histogram Equalization (CLAHE) to address the problem of underwater image degradation, and designed a real-time detection algorithm for underwater fish based on improved You Only Look Once (YOLO) and transfer learning. Compared with the naive YOLO, this method generates better detection performance at complex scenes of small fishes.

To address the issues of image degradation and detection robustness, Li *et al.* [40] introduced a frequency-enhanced detection model tailored for complex underwater scenes, which integrates high-frequency detail completion in the frequency domain with adaptive sharpening of spatial domain blur, achieving feature restoration for blurred images and thus improving recognition accuracy. Joint enhancement-detection optimization represents an integrated manner, where an enhancement module is co-designed/co-trained with a detector for optimal detection inputs. Liu *et al.* [41] proposed a plug-and-play Underwater joint image enhancement Module (UnitModule), which is jointly trained with the detector without requiring additional datasets, and enhanced input images to improve detector perception for object features. Chen *et al.* [42] presented SFDet, a small-object detection method for underwater images that integrates an enhancement network and extracts composite features from both spatial and frequency domains for enhancing detection accuracy of small objects. These preprocessing and augmentation strategies mitigates inherent challenges of underwater imaging and enhance model generalization ability to some extent, but fails in distribution bias and detail loss. that underwater image enhancement and underwater object detection have different objectives: the former aims to improve low-level visual quality, whereas the latter focuses on object recognition and localization. Better visual quality may not lead to better detection performance [43, 44, 45]. In underwater scenes, image enhancement alters visual characteristics relevant to object detection, but its effects vary across enhancement methods and detectors, thus the effectiveness of enhancement results should be evaluated not only on the visual quality but also on the actual contribution to downstream detection performance.

Multi-scale feature fusion. To address the coexistence of small objects with medium-scale and large-scale objects under water scenes, researchers have refined Feature Pyramid Network (FPN) to improve detection performance, which focus on fusing features across different network layers. Qi *et al.* [49] introduced a two-stage underwater small target detection network, where a deformable convolutional pyramid is proposed to handle the problems of deformation, occlusion, and various object sizes, strengthening the network to detect underwater small targets in complex environments. Du *et al.* [50] designed a weighted Bidirectional Feature Pyramid Network (BiFPN), which fuses multi-scale features by learnable weights and enhances network detection ability of varying underwater objects sizes. Wei *et al.* [51] presented a Hybrid Bridging Feature Pyramid Network (HBFPN), which merges deep semantic features with shallow detail features to improve feature extraction for blurry small underwater objects. Liu *et al.* [52] adopted a Generalized Feature Pyramid Network (GFPN), utilizing cross-layer connections and an adaptive feature selection mechanism to enhanced feature extraction pertinent to underwater small targets. During the multi-scale fusion, Lei *et al.* [53] used a Feature Fusion Module (FFM) to address unbalanced semantic distribution and local detail loss, thereby enhancing the model accuracy of detecting underwater small targets. To enhance multi-scale fusion efficacy, adaptive and dynamic feature selection mechanisms have been widely used. Some studies employs dynamic channel selection blocks to optimize the feature use from varying receptive fields [54], or utilized adaptive multi-scale fusion modules to integrate contextual information [55]. Other work are dynamic feature fusion algorithms, using cascaded dynamic modules and local-global fusion blocks to enhance feature robustness for underwater targets [56]. In addition, dilated convolution and its variants are employed to expand the network's receptive field, thereby enhancing feature extraction capabilities for multi-scale underwater targets. By designing a multi-rate dilated convolution module, multi-scale features can be captured in parallel without downsampling, addressing the scale diversity of underwater targets [57]. The switchable dilated convolution module dynamically selects most appropriate receptive field size based on input content, and adapts to feature extraction requirements of different-scale targets [58]. Table 4 summarizes improvement strategies of multi-scale fusion.

Table 4 Comparison of multi-scale feature fusion strategies

Strategy	Ref.	Advantages	Limitations
Pyramid Structure	[50, 59]	Cross-layer interaction	Computational burden
	[60, 61, 62]	Dynamic attention	Extra parameters
	[63, 64]	Lightweight design	Accuracy degradation
Receptive Field Expansion	[57, 58]	Multi-scale context	Rate tuning
	[47, 65]	Adaptive deformation	Training instability
	[66]	Dynamic field adjustment	Complex module
	[67]	Long-range dependencies	High resources

Attention mechanisms introduction. In underwater small target detection, attention mechanisms have been widely adopted to address issues of target blurring, small scales, and complex backgrounds, which could enhance model perception and discriminative capabilities through adaptive weighting of features. Fig. 11 shows four main ways of existing methods for enhancing features: Channel attention recalibrates weights of feature channels to strengthen key feature expression, including the Squeeze-and-Excitation (SE) module [68, 69], Effective Squeeze-and-Excitation (ESE) mechanism [60], Efficient Channel Attention (ECA) structure [47], and application of channel attention in feature fusion [51, 70, 71]. Spatial attention focuses on spatial location features of targets, represented by coordinate attention [60, 72, 73, 74], convolutional attention mechanisms [75], and multi-scale spatial attention modules [76, 77]. Channel-spatial hybrid attention considers both channel and spatial dimensions simultaneously, exemplified by structures such as the Convolutional Block Attention Module (CBAM) [48, 52, 78], dual attention combinations [51, 67], Efficient Multi-scale Attention (EMA) [52, 64, 71, 79, 80], as well as SimAM [81] and Aggregated Multi-Scale Attention Module (AMSA) [82]. Moreover, self-attention mechanisms leverage Transformer architectures to capture long-range dependencies, such as single-/multi-head self-attentions [77, 83], hybrid Transformer backbones [67], MobileViT [73], and frequency-domain self-attentions [40]. These attention modules are embedded into backbone networks [60, 61, 68, 84], feature pyramids [48, 50, 51, 85, 86], or detection heads [65, 67, 72, 81], thus improving detecting performance of underwater small objects in terms of feature enhancement, noise robustness, scale adaptability, and global context modeling.

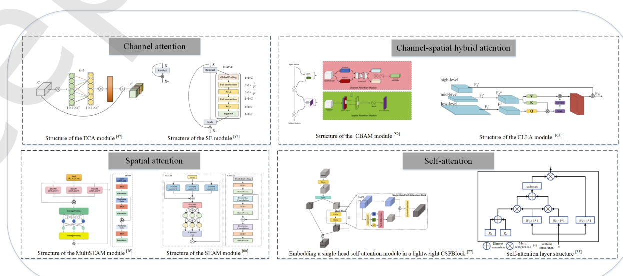

Fig. 11 Examples of attention mechanism introduction.

Loss function optimization. Table 5 summarizes different strategies of optimizing loss functions in underwater small object detection. Early works focused on fine-tuning losses for individual tasks, and current trends involve constructing composite loss functions to optimize multiple dimensions of detection task. By performing the weighted fusion of IoU-variant losses targeting localization accuracy, classification losses address sample imbalance, and auxiliary losses are tailored for specific scenarios, such as density-aware losses for crowded targets and perception losses for occluded objects. This design [40, 67, 70, 88] enables models to learn in a more balanced manner, and achieves more robust performance in complex underwater environments. Geometric metrics based on Intersection over Union (IoU) are highly sensitive to positional deviations in small objects, where minor shifts can lead to the gradient vanishing. Thereby, the research shifts from geometric metrics toward probabilistic distribution metrics. The Normalized Wasserstein Distance (NWD) [46, 51, 84, 87] model is proposed with bounding boxes as two-dimensional Gaussian distributions, and the distance between boxes is measured by computing distribution similarity. This approach mitigates training instability by annotation noise or slight localization inaccuracies in small objects, and has become a key technique to enhance the detection accuracy of small objects. As application scenarios become specialized, loss function design reflects tailored adaptations to underwater environments. Density-aware losses are designed to detect dense schools of small fish [89], blur-adaptive losses are introduced to address image blurring [40], and regression losses focused on hard samples are employed to improve detection rate of occluded targets [60]. This strategy of scenario-defined loss enables algorithms to address practical challenges encountered during real-world deployment.

Lightweight design. Various lightweight strategies have been developed to meet computational and memory constraints of underwater embedded devices. Table 6 lists lightweight design strategies for underwater small object detection. Among them, redesigning backbone network is one of most effective approaches. Zhang *et al.* [72] replaced the naive Darknet-53 backbone in YOLOv8s with FasterNet-T0, improving detection accuracy while reducing parameters. At the module level, re-parameterization modules and efficient convolution designs are widely adopted. Zhou *et al.* [92] enhanced feature extraction for small and dense objects by integrating RFAConv modules and optimizing dual structure, reducing computational costs.

In complex underwater small object detection, a single improvement strategy is insufficient to overcome combined challenges of environmental interference, scarce features, and efficiency constraints, thereby failing to achieve comprehensive performance improvement. Hence, a core mind is adopted to integrate multiple complementary strategies in algorithm design.

5 Challenges & Prospects

Underwater small object detection technique has made significant progress in recent years, and has been widely applied in the fields of marine organism detection, mineral resource exploration, etc. However, this technology encounters numerous challenges at complex underwater environments, and these issues can be summarized into three aspects below:

***<u>Datasets</u>*: class imbalance and limited scenario.** Existing underwater image datasets fall short of meeting the training and generalization requirements of small object detection models. In terms of class distribution, easily collected objects such as sea urchins account for over 50% of samples, while other small biological categories are underrepresented [27, 28, 29]. This imbalance biases model training toward dominant classes, reducing detection accuracy for scarce categories. Regarding scenario coverage, most underwater images come from stable, clear environments such as coastal and shallow seas, with relatively few samples from special scenarios such as the deep sea, inland rivers, and turbid waters. Consequently, model generalization drops significantly in these scarce environments [97]. Moreover, due to the small pixel size and blurred features of underwater small objects, the manual annotation is prone to bounding-box misalignment, missed annotations, and duplicate labeling. The differences in annotators' judgments of object boundaries may introduce annotation noise, affecting both model training and performance evaluation. For underwater small objects, minor boundary deviations can cause noticeable IoU changes, making annotation quality particularly important for reliable training and evaluation.

***<u>Algorithms</u>*: lightweight deployment versus high accuracy.** A major challenge in algorithm design is a trade-off between lightweight model and detection accuracy [7]. Both low pixel proportion and weak features of underwater small objects demand high feature extraction capability, but adding modules for higher accuracy increases model complexity and parameter count, raising computational burden beyond limited resources and power budgets of edge devices like underwater robots. Most of existing algorithms are validated in lab settings on standardized datasets, lacking real-world deployment testing. Thus, achieving high detection accuracy while maintaining lightweight model under hardware constraints is still a core challenge. Besides, the degraded issues of color distortion, low contrast, and blur in underwater images weaken the robustness and adaptability of existing detection models, thereby leading to unstable detection accuracy.

***<u>Metrics</u>*: lacking specific evaluation of small objects.** The performance evaluation of underwater small object detection algorithms focus on two aspects: accuracy and speed. IoU measures localization accuracy by calculating the overlap between predicted and ground-truth bounding boxes. Precision reflects the proportion of true positives among all detections predicted as positive, indicating model's ability to avoid false alarms. Recall represents the proportion of actual positives that are correctly identified, measuring detection completeness. Average Precision (AP) synthesizes precision across different recall levels to evaluate performance for a single category. Mean Average Precision (mAP) averages AP values across all categories to assess overall performance in multi-class detection. Among them, the mAP metric is mostly used to evaluate the performance of underwater small object detection, which averages contributions across all object sizes. As a result, a model may maintain a high mAP score even if it performs well on medium and large objects while completely failing on small ones [98]. Hence, existing metrics do not reflect small object detection capability adequately. Further, due to the characteristics of low pixel proportion and dense overlapping, underwater small objects are prone to missed detections—not sufficiently measured by current metrics.

In response to the above challenges, future development of underwater small object detection technique can be predicted to be categorized into four key areas:

Dataset expansion and standardization. Efforts should center on building underwater small object datasets that cover diverse scenarios, and image collection should be extended to deep sea, inland rivers, and other extreme environments, forming a standardized data resource that includes varying depths, water qualities, and lighting conditions, to ultimately strengthen the model generalization in real underwater settings. Meanwhile, future works will focus on achieving multiclass balance and standardizing the annotation process, and collaboration with marine research teams can help increase samples of rare small biological categories. Semi- and weakly-supervised labeling techniques can be adopted to reduce annotation costs [99]. Moreover, establishing an unified criteria for determining small object boundaries will minimize annotator subjectivity and improve the consistency, thereby enhancing the effectiveness of model training.

Algorithm lightweight and deployment. Lightweight design and high accuracy will be synergistically improved for underwater small object detection. It can be achieved via designing dedicated lightweight modules for feature extraction

Table 5 Summary of loss function optimization schemes for underwater small object detection

Ref.	Loss function	Advantages	Applications
[50]	Focaler-IoU	Linear interval mapping for small object sensitivity	Blurred and occluded small objects
[60]	WIoUv2	Improve occluded object accuracy	Occluded small objects
[68]	EIoU	Small object feature extraction with reduced complexity	Underwater small objects
[85]	Shape-IoU	Geometry-regression relationship	Small objects with varying shapes
[90]	Inner-CIoU	Generalization to low-visibility environments	Low-visibility small objects
[46]	NWD	Localization sensitivity reduction	Small object matching
[58]	Wise-IoU + NWD	Dual metric for multi-scale sample imbalance	Multi-scale sample imbalance
[91]	Wasserstein Loss	Positional sensitivity reduction	Micro-organism detection
[40]	Classification + Regression + Blur-Aware Loss	Convergence improvement	Blurred small objects
[61]	CIoUv2 + Focal Loss v2	Localization and rare sample classification balance	Scarce biological specimen
[70]	Classification + Regression + Density + Occlusion Loss	Reduce missed detections and duplicate boxes	Dense small objects

Table 6 Summary of lightweight design strategies for underwater small object detection

Ref.	Lightweight design	Advantages	Parameters (M)
[47]	BiFPN with optimization	Multi-scale feature fusion with fewer parameters	2.10
[51]	Ref-Dilated Conv block	Tradeoff between feature extraction and efficiency	1.80
[71]	P-C2F + Lightweight FPN	Reduced redundant computation	1.64
[73]	MobileViT backbone	Global feature representation with fewer parameters	4.51
[76]	RFAConv + Dual_C2f	Dynamic feature extraction for lower computation	3.00
[79]	FasterBlock in C2f	Reduced computation in critical modules	8.41
[82]	Simplified GSConv	Minimal parameters and computation	3.00
[93]	GhostNet + LAMP pruning	Reduced model parameters and size	1.86
[94]	Context-guided Block	Feature extraction with less model complexity	14.86
[95]	C2f-G Module	Lower computation with efficient structure	18.00
[96]	Depth-wise separable Conv	Reduced model parameters	10.73

of underwater small objects, combined with advanced techniques of model compression such as knowledge distillations [100] and prunings [101], to strike an optimal tradeoff between model size and detection performance. Introducing adaptive anti-interference mechanisms into network structures will enhance the robustness against underwater degradations of color distortions and blurs. Additionally, conducting real-scene deployment by optimized algorithms on underwater robotics and testing system performance across actual environments will continuously improve model's real-time capability and stability in dynamic underwater conditions.

Metric development of small objects. To accurately assess the performance of underwater small object detection, dedicated metrics for small targets should be designed. Similar to the standard COCO evaluation protocol [102], objects can be strictly divided into small, medium, and large scales based on pixel area, with average precision calculated independently for each scale. It is worth considering that the average precision for small objects should serve as a core metric, reflecting model's detection capability on small scale objects. Moreover, a dedicated small object miss-detection rate can be established by analyzing the proportion of undetected small objects among ground-truth annotations, thus revealing perceptual blind spots of the model[103].

Multi-modal foundation models. Beyond underwater optical images, underwater environments provide complementary data, such as acoustic sonar, depth information, water parameters, and so on. Foundation models with cross-modal representation and alignment capabilities, will offer a new pathway to integrate multi-modal data. These latest models can construct a unified semantic representation space, combining texture details from optical images with spatial structures from sonar data, to enable synergistic enhancement of cross-modal features [104]. Recent studies also demonstrate the potential of foundation models to exploit region-level priors for underwater semantic understanding [105]. This manner promises sustained and stable detection of small objects in complex underwater conditions.

6 Conclusion

In order to address the gap of lacking reviews on underwater small object detection, this paper delivers a comprehensive survey of small object detection in underwater images. First, we categorize the core problems of underwater small object detection into inherent object characteristics and underwater image degradation, then we sort out and detail 16 existing typical public datasets. Moreover, we summarize the methodology evolution from traditional machine learning methods to deep learning-based approaches, and investigate an in-depth analysis on deep learning-based methods with five types of improvement strategies. Furthermore, we point out current challenges faced by underwater small object detection, including imbalanced datasets, lightweight accuracy trade-off, and unsuitable evaluation metrics. Although encountering the above challenges, underwater small object detection remains promising, and its future directions will focus on constructing comprehensive datasets, developing accurate and lightweight detection algorithms, dedicating underwater small object evaluation metrics, and exploring multi-modal foundation models for underwater object detections. In the end, this review will be serve as a valuable reference for subsequent study, and we believe that underwater small object detection will be a pivotal technology of advancing the fields of marine science and engineering.

7 Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grants 62171252, 61701245, 62501357; in part by the Fundamental Research Funds for the Central Universities under Grant 00007764.

References

[1] Z. R. Chen and Z. H. Wang. Research on underwater target detection using side-scan sonar and multibeam sounding system. Hydrographic Surveying and Charting. 33(4), 51-54, 2013.

[2] K. M. Dunlop, E. Marian Scott, D. Parsons and D. M. Bailey. Do agonistic behaviours bias baited remote underwater video surveys of fish. Marine Ecology. 36: 810-818, 2015.

[3] D. K. Rout, M. Kapoor, B. N. Subudhi, V. Thangaraj, V. Jakhetiya and A. Bansal. Underwater visual surveillance: A comprehensive survey. Ocean Engineering. 309: 118367, 2024.

[4] W. Wei, Y. Cheng, J. He and X. Zhu. A review of small object detection based on deep learning. Neural Computer and Application. 36: 6283-6303, 2024.

[5] S. Xu, M. Zhang, W. Song, H. Mei, Q. He and A. Liotta. A systematic review and analysis of deep learning-based underwater object detection. Neurocomputing. 527: 204-232, 2023.

[6] J. N. Dong, M. Y. Yang, R. Z. Zhuo and L. P. Cai. A review of underwater image target detection datasets and detection algorithms. Journal of Ocean Technology. 41(5): 60-72, 2022.

[7] M. Muhamad and X. Li. Comprehensive review of deep learning-based tiny object detection: challenges, strategies, and future directions. Knowledge and Information Systems. 1: 1-89, 2025.

[8] Z. Moorton, K. Mistry, R. Strachan and S. Hu. Up-to-Date Scoping Review of Object Detection Methods for Macro Marine Debris. Journal of Marine Science and Engineering. 13(8): 1590, 2025.

[9] U. Nawaz, M. Anees-ur-Rahaman and Z. Saeed. A Survey of Deep Learning Approaches for the Monitoring and Classification of Seagrass. Ocean Science Journal. 60(2): 19, 2025.

[10] C. Y. Lu, F. G. Jia and L. M. Yu. A Review of Underwater Image Target Detection Research. Digital Ocean & Underwater Warfare. 6(1): 34-40, 2023.

[11] J. B. Zhao and B. S. Du. Development of small object detection technology based on deep learning. Electronics Optics & Control. 30(2): 63-70, 74, 2023.

[12] S. Fayaz, S. A. Parah and G. J. Qureshi. Underwater object detection: architectures and algorithms–a comprehensive review. Multimedia Tools and Applications. 81(15): 20871-20916, 2022.

[13] D. Gomes, A. F. M. S. Saif and D. Nandi. Robust Underwater Object Detection with Autonomous Underwater Vehicle: A Comprehensive Study. Proceedings of the International Conference on Computing Advancements (ICCA). New York, NY, USA: Association for Computing Machinery, Article 17, 1-10, 2020.

[14] M. Moniruzzaman, S. M. S. Islam, M. Bennamoun and P. Lavery. Deep Learning on Underwater Marine Object Detection: A Survey. Advanced Concepts for Intelligent Vision Systems (ACIVS). Lecture Notes in Computer Science, vol. 10617, Cham: Springer, 2017.

[15] A. Torralba, R. Fergus and W. T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 30(11): 1958-1970, 2008.

[16] C. Chen, M. Y. Liu, O. Tuzel and J. Xiao. R-CNN for small object detection. Proceeding of Asian Conference on Computer Vision (ACCV). Cham: Springer, 214-230, 2016.

[17] Y. S. Tsai, C. T. Tsai and J. H. Huang. Multi-scale Detection of Underwater Objects Using Attention Mechanisms and Normalized Wasserstein Distance Loss. The Journal of Supercomputing. 81(6): 785, 2025.

[18] R. Jing, W. Zhang, Y. Li, W. Li and Y. Liu. Feature aggregation network for small object detection. Expert Systems with Applications. 255: 124686, 2024.

[19] G. Dong, W. C. Xie, X. L. Huang, Y. T. Qiao and Q. Mao. Review of deep learning small object detection algorithms. Computer Engineering and Applications. 59(11): 16-27, 2023.

[20] Z. Zhou and Y. Zhu. Libra-SOD: balanced label assignment for small object detection. Knowledge Based System. 302: 112353, 2024.

[21] W. Chen, F. Yuan, E. Cheng and W. Lin. Subjective and objective quality evaluation of sonar images for underwater acoustic transmission. IEEE International Conference on Image Processing (ICIP). pp. 176-180, 2017.

[22] F. Bonin, A. Burguera and G. Oliver. Imaging systems for advanced underwater vehicles. Journal of Maritime Research. 8(1): 65-86, 2011.

[23] S. Wang, K. Ma, H. Yeganeh, Z. Wang and W. Lin. A patch-structure representation method for quality assessment of contrast changed images. IEEE Signal Processing Letters. 22(12): 2387-2390, 2015.

[24] G. Cutter, K. Stierhoff and J. Zeng. Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: labeled fishes in the wild. IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW). 57-62, 2015.

[25] M. Pedersen, J. B. Haurum, R. Gade and T. B. Moeslund. Detection of marine animals in a new underwater dataset with varying visibility. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 18-26, 2019.

[26] M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. S. Enan and J. Sattar. Semantic Segmentation of Underwater Imagery: Dataset and Benchmark. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1769-1776, 2020.

[27] C. W. Liu, Z. H. Wang, S. J. Wang, T. Tang, Y. L. Tao and C. F. Yang. A new dataset, poisson gan and aquanet for underwater object grabbing. IEEE Transactions on Circuits and Systems for Video Technology. 32(5): 2831-2844, 2021.

[28] C. Liu, H. Li, S. Wang, M. Zhu, D. Wang, X. Fan and Z. Wang. A dataset and benchmark of underwater object detection for robotpicking. IEEE International Conference on Multimedia and Expo Workshops (ICMEW). pp. 1-6, 2021.

[29] L. Jiang, Y. Wang, Q. Jia, S. Xu, Y. Liu, X. Fan, H. Li, R. Liu, X. Xue and R. Wang. Underwater species detection using channel sharpening attention. Proceedings of the 29th ACM International Conference on Multimedia (ACM MM). pp. 4259-4267, 2021.

[30] C. Fu, R. Liu, X. Fan, P. Chen, H. Fu, W. Yuan, M. Zhu and Z. Luo. Rethinking general underwater object detection: Datasets, challenges, and solutions. Neurocomputing. 517: 243–256, 2023.

[31] A. Đuraš, B. J. Wolf, A. Ilioudi, I. Palunko and B. De Schutter. A dataset for detection and segmentation of underwater marine debris in shallow waters. Scientific Data. 11: 921, 2024.

[32] R. Mukherjee, S. Singh, J. McWilliams and J. Sattar. The Common Objects Underwater (COU) Dataset for Robust Underwater Object Detection.IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hangzhou, China, pp. 18597-18603, 2025.

[33] N. T. Retta, A. Parulkar, A. Chowdhury, A. Venkatesh, Y. Seelam, N. Reddy, S. Jayaraj and K. Chandrasekaran. DebrisVision: Bridging the synthetic-to-real gap for enhanced underwater debris analysis. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp. 2105-2114, 2025.

[34] M. Fatan, M. R. Daliri, and A. M. Shahri. Underwater cable detection in the images using edge classification based on texture information. Measurement. 91: 309-317, 2016.

[35] S. Villon, M. Chaumont, G. Subsol, S. Villéger, T. Claverie, and D. Mouillot. Coral Reef Fish Detection and Recognition in Underwater Videos by Supervised Machine Learning: Comparison Between Deep Learning and HOG+SVM Methods.Advanced Concepts for Intelligent Vision Systems (ACIVS 2016). pp. 160-171, 2016.

[36] C. S. Tan, P. Y. Lau, and T. J. Low. Macroalgae Recognition Based on Histogram Oriented Gradient. 9th International Conference on Robotic, Vision, Signal Processing and Power Applications. pp. 257-266, 2017.

[37] A. Mathias, S. Dhanalakshmi, R. Kumar and R. Narayanamoorthi. Underwater object detection based on bi-dimensional empirical mode decomposition and Gaussian Mixture Model approach. Ecological Informatics. 66: 101469, 2021.

[38] H. Luo, H. B. Ruan and D. W. Tu. Research on small sample target detection for underwater robot. Robotic Intelligence and Automation. 44(2): 229-241, 2024.

[39] Q. Z. Li, Y. B. Li and J. Niu. Real-time detection algorithm of underwater fish targets based on improved YOLO and transfer learning. Pattern Recognition and Artificial Intelligence. 32(3): 193-203, 2019.

[40] Y. Li, J. Deng, Z. Zou, Q. Yang, Q. Ren and Y. Pan. AF-DETR: A Dual-Stream Frequency Enhancement and Blur-Adaptive Transformer for Underwater Object Detection. Research Square. 2025.

[41] Z. Y. Liu, B. Wang, Y. Li, J. X. He and Y. F. Li. UnitModule: A lightweight joint image enhancement module for underwater object detection. Pattern Recognition. 151: 110435, 2024.

[42] D. Z. Chen and G. Gou. SFDet: spatial to frequency attention for small-object detection in underwater images. Journal of Electronic Imaging. 33(2): 023057, 2024.

[43] H. Wang, S. Sun, X. Bai, J. Wang and P. Ren. A Reinforcement Learning Paradigm of Configuring Visual Enhancement for Object Detection in Underwater Scenes. IEEE Journal of Oceanic Engineering. 48(2): 443-461, 2023.

[44] Y. Wang, J. Guo, W. He, H. Gao, H. Yue, Z. Zhang and C. Li. Is Underwater Image Enhancement All Object Detectors Need? IEEE Journal of Oceanic Engineering. 49(2): 606-621, 2024.

[45] H. Li, L. Li, H. Wang, W. Zhang, J. Liu and P. Ren. Joint Underwater Image Enhancement and Captioning Through Multisupervised Chained Task Learning. IEEE Transactions on Geoscience and Remote Sensing. 64: 4209622, 2026.

[46] W. W. Chen, T. T. Zhuang, Y. F. Zhang, T. Mei and X. Y. Tang. YOLO-UOD: An underwater small object detector via improved efficient layer aggregation network. IET Image Processing. 18(9): 2490-2505, 2024.

[47] X. H. Wang, X. Y. Song, Z. Li and H. Q. Wang. YOLO-DBS: Efficient Target Detection in Complex Underwater Scene Images Based on Improved YOLOv8. Journal of Ocean University of China. 24(4): 979-992, 2025.

[48] P. Z. Liu, W. B. Qian and Y. L. Wang. YWnet: A convolutional block attention-based fusion deep learning method for complex underwater small target detection. Ecological Informatics. 79: 102401, 2024.

[49] S. Qi, J. Du, M. Wu, H. Yi, L. Tang, T. Qian and X. Wang. Underwater Small Target Detection Based on Deformable Convolutional Pyramid. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, Singapore, pp. 2784-2788, 2022.

[50] R. S. Du, Z. S. Wang, L. D. Meng and Y. G. Jing. Underwater small target detection based on self-attention and improved pyramid. Journal of Shenzhen University Science & Engineering. 42(4): 428-436, 2025.

[51] Y. F. Wei, J. Tao, W. J. Wu, D. H. Yuan and S. Z. Hou. RHS-YOLOv8: A Lightweight Underwater Small Object Detection Algorithm Based on Improved YOLOv8. Applied Sciences. 15(7): 3778, 2025.

[52] L. L. Liu, C. X. Chu, C. C. Chen and S. D. Huang. MarineYOLO: Innovative deep learning method for small target detection in underwater environments. Alexandria Engineering Journal. 104: 423-433, 2024.

[53] L. Deng, S. J. Luo, C. H. He, H. P. Xiao and H. Wu. Underwater small and occlusion object detection with feature fusion and global context decoupling head-based YOLO. Multimedia Systems. 30: 208, 2024.

[54] W. J. Ouyang, Y. H. Wei and D. X. Wang. YOLOX-DC: A Small Target Detection Network up to Underwater Scenes. IEEE OCEANS. 2022.

[55] L. S. Luo, Q. Gao, Q. Liu, X. Li and X. Yu. Multi-scale adaptive small and overlapping target detection in underwater images. Journal of Supercomputing. 81: 1474, 2025.

[56] W. Song, Z. Y. Chen, Z. Y. Jiang, J. L. Deng, H. P. Cui, X. B. He and J. X. Xia. Dynamic feature fusion algorithm for multi-scale biological target detection in underwater environments. Digital Signal Processing. 170: 105786, 2026.

[57] Y. Z. Chen, Y. Z. Huang and J. H. Zhang. Underwater Small Object Detection Based on Multi-Rate Dilated Convolution. Computer Engineering. 49(6): 257-264, 2023.

[58] M. H. Zhang, J. P. Huang, W. Song, Q. H. Xiao and D. F. Zhao. Multi-scale Object Detection Method for Underwater Organisms under Unbalanced Samples Based on UMS-YOLO v7. Transactions of the Chinese Society for Agricultural Machinery. 56(1): 388-396, 409, 2025.

[59] M. Y. Qiao, J. K. Shi, B. F. Li, Y. Zhao and Y. Q. Shi. Enhanced FPN underwater small target detection with improved loss function. Journal of Computer-Aided Design & Computer Graphics. 35(4): 525-537, 2023.

[60] J. C. Li, Y. L. Han and C. S. Dong. YOLO-U: Underwater target detection algorithm based on structural reparameterization and dual attention mechanism. Journal of Shanghai Ocean University. 34(3): 696-706, 2025.

[61] D. X. Wang, Y. He and H. C. Yuan. Underwater biological target detection method based on YOLOv8-BAN model. Jiangsu Journal of Agricultural Sciences. 1: 101-111, 2025.

[62] J. X. Gao, Y. H. Zhang, X. Geng, H. T. Hao and U. A. Bhatti. PE-Transformer: Path enhanced transformer for improving underwater object detection. Expert Systems with Application. 246(C): 123253, 2024.

[63] M. Liu, Y. Wu, R. Li and C. Lin. LFN-YOLO: precision underwater small object detection via a lightweight reparameterized approach. Frontiers in Marine Science. 11: 1513740, 2025.

[64] Z. F. Ye, X. Peng, D. K. Li and F. Shi. YOLOv11-MSE: A Multi-Scale Dilated Attention-Enhanced Lightweight Network for Efficient Real-Time Underwater Target Detection. Journal of Marine Science and Engineering. 13(10): 1843, 2025.

[65] C. T. Dang, H. Sato and M. Kubo. DWP2-YOLO11: Deformable Convolution and Multi-Scale Head for Underwater Object Detection. 8th International Conference on Control, Robotics and Informatics (ICCRI). Danang, Vietnam, pp. 21-25, 2025.

[66] Y. J. Guo, X. F. Zhao, Z. M. Zhong and X. M. Zhong. GAI-YOLO: an underwater target detection algorithm based on enhanced feature representation and adaptive receptive field. Journal of Supercomputing. 81: 1516, 2025.

[67] G. Chen, Z. Mao, K. Wang and J. Shen. HTDet: A Hybrid Transformer-Based Approach for Underwater Small Object Detection. Remote Sensing. 15(4): 1076, 2023.

[68] W. G. Yi and B. Wang. Research on Underwater Small Target Detection Algorithm Based on Improved YOLOv7. IEEE Access. 11: 66818-66827, 2023.

[69] J. F. Li, Y. W. Zhu, M. X. Chen, Y. L. Wang and Z. Q. Zhou. Research on Underwater Small Target Detection Algorithm Based on Improved YOLOv3. 16th IEEE International Conference on Signal Processing (ICSP). pp. 76-80, 2022.

[70] J. Wang, Y. Li, J. Wang and Y. Li. An underwater dense small object detection model based on YOLOv5-CFDSDSE. Electronics. 12(15): 3231, 2023.

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

[72] M. Zhang, Z. Wang, W. Song, D. Zhao and H. Zhao. Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network. Applied Sciences. 14(3): 1095, 2024.

[73] Y. Sun, W. X. Zheng, X. Du and Z. P. Yan. Underwater Small Target Detection Based on YOLOX Combined with MobileViT and Double Coordinate Attention. Journal of Marine Science and Engineering. 11(6): 1178, 2023.

[74] M. F. Huang, W. M. Huang, Y. N. Xiao, Y. Y. Wen and Z. R. Deng. YOLOV5 Underwater Target Detection Algorithm Based on Multi-Aggregation and Multi-Scale. Computer and Digital Engineering. 52 (11): 3197-3201, 3309, 2024.

[75] L. Chen, D. Q. Huang, H. J. Shi and Z. F. Wang. Underwater small target detection algorithm based on multi-level feature aliasing fusion. Journal of Detection & Control. 44(5): 77-82, 2022.

[76] X. C. Gao, X. Y. Bai and P. X. Liu. YOLO-RDM: A Lightweight and High-Precision Method for Underwater Small Target Detection. 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD). pp. 1020-1025, 2025.

[77] J. Peng, Z. Zhou and H. Wang. Lightweight real-time detection transformer with Single-Head Self-Attention and Feature Selection for underwater object detection. Journal of Real-Time Image Processing. 22: 173, 2025.

[78] Z. C. Li, H. J. Xie, J. Y. Feng, Z. B. Wang and Z. Z. Yuan. YOLOv7-PE: A Precise and Efficient Enhancement of YOLOv7 for Underwater Target Detection. IEEE Access. 12: 133937-133951, 2024.

[79] D. Y. Cao, H. Yang, C. Guo, Y. Y. Cheng, W. T. Zhang, M. J. Shi and X. Q. Xia. ESC-YOLOv8-seg: A real-time non-destructive detection framework for small-target surface anomalies in zebrafish underwater monitoring. Ecological Informatics. 92: 103473, 2025.

[80] Y. Zhang and X. Zhang. YOLO-AES A Lightweight Model for Real-Time Underwater Object Detection. 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Wenzhou, China. pp. 611-615, 2024.

[81] D. H. Lu, J. X. Yi and J. Wang. Enhanced YOLOv7 for Improved Underwater Target Detection. Journal of Marine Science and Engineering. 12(7): 1127, 2024.

[82] W. K. Sun, X. Q. Liu, J. Hao, Q. Y. Yao, H. L. Xi, Y. W. Wu and Z. Y. Xing. AGS-YOLO: An Efficient Underwater Small-Object Detection Network for Low-Resource Environments. Journal of Marine Science and Engineering. 13(8): 1465, 2025.

[83] X. T. Qi and H. C. Yuan. Small target biological detection based on self-attention mechanism and improved YOLOv5s. Journal of Hunan Agricultural University: Natural Sciences. 50(3): 108-114, 2024.

[84] S. M. Qu, C. Cui, J. L. Duan, Y. Y. Lu and Z. L. Pang. Underwater small target detection under YOLOv8-LA model. Scientific Reports. 14(1): 16108, 2024.

[85] Z. W. Zheng and W. W. Yu. RG-YOLO: multi-scale feature learning for underwater target detection. Multimedia Systems. 31: 26, 2025.

[86] H. Liu, C. M. Luo, Z. Q. Huang and Z. C. Dou. Underwater small object detection method enhanced by multi-level feature fusion. Proceedings of the International Conference on Image, Signal Processing, and Pattern Recognition (ISPP). Guangzhou, China. pp. 1318043, 2024.

[87] M. Zhao, H. B. Zhou and X. Li. YOLOv7-SN: Underwater Target Detection Algorithm Based on Improved YOLOv7. Symmetry-Basel. 16(5): 514, 2024.

[88] J. Y. Shi and W. J. Wu. Srp-uod: Multi-branch hybrid network framework based on structural re-parameterization for underwater small object detection. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024.

[89] X. J. Liu, Y. Liu and Z. L. Wei. Underwater Dense Small Object Fish Detection Based on CCA-YOLOv8. Journal of Jiamusi University (Natural Science Edition). 43(8): 1-4, 2025.

[90] W. G. Yi, J. W. Yang and L. W. Yan. Research on Underwater Small Target Detection Technology Based on Single-Stage USSTD-YOLOv8n. IEEE Access. 12: 69633-69641, 2024.

[91] L. Liu, J. F. Wu, H. Y. Zhao, H. Kong, T. Zheng, B. Y. Qu and H. Yu. Fish-Finder: A robust small target detection method for aquaculture fish in low-quality underwater images. Journal of Fish Biology. 106(3): 908-920, 2025.

[92] Z. Y. Zhou, C. X. Ma, L. S. Yang, Z. M. Zhong and W. B. Hu. Lightweight improved YOLOv8 underwater target detection model. Electronic Measurement Technology. 47(19): 181-189, 2024.

[93] Y. C. Wang, G. J. Hou, J. Q. Ma and Y. H. Cui. Lightweight underwater small target detection method based on YOLO12. Digital Ocean & Underwater Warfare. 8(4): 417-423, 2025.

[94] Z. Dong, Q. Yang, H. L. Chen and D. X. Gao. Deep learning based lightweight real-time detection framework for small target in complex underwater environments. Multimedia Systems. 32: 44, 2025.

[95] V. Sineglazov, M. Savchenko and V. Lytvynenko. Efficient Small and Overlapping Target Detection in Underwater Images. 4th International Workshop of IT-Professionals on Artificial Intelligence. pp. 16-29, 2024.

[96] M. Zhang, S. Xu, W. Song, Q. He and Q. Wei. Lightweight Underwater Object Detection Based on YOLO v4 and Multi-Scale Attentional Feature Fusion. Remote Sensing. 13(22): 4706, 2021.

[97] E. Nabahirwa, W. Song, M. Zhang, Y. Fang and Z. Ni. A Structured Review of Underwater Object Detection Challenges and Solutions: From Traditional to Large Vision Language Models. arXiv preprint. arXiv:2509.08490, 2025.

[98] J. Gilg, T. Teepe, F. Herzog and G. Rigoll. The Box Size Confidence Bias Harms Your Object Detector. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1471-1480, 2023.

[99] R. Rahali, T. P. Nguyen and V. Nguyen. Enhancing Marine Habitats Detection: A Comparative Study of Semi-Supervised Learning Methods. Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP). 3: 233-244, 2025.

[100] G. Hinton, O. Vinyals and J. Dean. Distilling the knowledge in a neural network. NIPS Deep Learning and Representation Learning Workshop. 2015.

[101] S. Han, H. Mao and W. J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. International Conference on Learning Representations (ICLR). 2016.

[102] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L. Zitnick. Microsoft COCO: Common objects in context. European Conference on Computer Vision (ECCV). pp. 740-755, 2014.

[103] A. Sobti, V. Mavi, M. Balakrishnan and C. Arora. VmAP: A Fair Metric for Video Object Detection. Proceedings of the 29th ACM International Conference on Multimedia (ACM MM). pp. 2224-2232, 2021.

[104] F. Yu, F. Xiao, C. Li, E. Cheng and F. Yuan. AO-UOD: A Novel Paradigm for Underwater Object Detection Using Acousto–Optic Fusion. IEEE Journal of Oceanic Engineering. 50(2): 919-940, 2025.

[105] H. Li, L. Li, H. Wang, W. Zhang and P. Ren. Large Foundation Model Empowered Region-aware Underwater Image Captioning. International Journal of Computer Vision. 134(2): 66, 2026.