

Review

Recent Progress in Ocean Intelligent Perception and Image Processing and the Impacts of Nonlinear Noise

Huayu Liu ¹, Ying Li ^{1,2,*} , Tao Qian ^{1,2} and Ye Tang ³ ¹ School of Design, Anhui Polytechnic University, Wuhu 241000, China; huayuliu@ahpu.edu.cn (H.L.); qiantao@ahpu.edu.cn (T.Q.)² Anhui Provincial Key Laboratory of Intelligent Technology and Design Culture on Philosophy and Social Sciences, Wuhu 241000, China³ Ocean Institute, Northwestern Polytechnical University, Suzhou 215000, China; tangye@nwpu.edu.cn

* Correspondence: liying@ahpu.edu.cn; Tel.: +86-188-9531-6694

Abstract: Deep learning network models are crucial in processing images acquired from optical, laser, and acoustic sensors in ocean intelligent perception and target detection. This work comprehensively reviews ocean intelligent perception and image processing technology, including ocean intelligent perception devices and image acquisition, image recognition and detection models, adaptive image processing processes, and coping methods for nonlinear noise interference. As the core tasks of ocean image processing, image recognition and detection network models are the research focus of this article. The focus is on the development of deep-learning network models for ocean image recognition and detection, such as SSD, R-CNN series, and YOLO series. The detailed analysis of the mathematical structure of the YOLO model and the differences between various versions, which determine the detection accuracy and inference speed, provides a deeper understanding. It also reviewed adaptive image processing processes and their critical support for ocean image recognition and detection, such as image annotation, feature enhancement, and image segmentation. Research and practical applications show that nonlinear noise significantly affects underwater image processing. When combined with image enhancement, data augmentation, and transfer learning methods, deep learning algorithms can be applied to effectively address the challenges of underwater image degradation and nonlinear noise interference. This work offers a unique perspective, highlighting the mathematical structure of the network model for ocean intelligent perception and image processing. It also discusses the benefits of DL-based denoising methods in signal–noise separation and noise suppression. With this unique perspective, this work is expected to inspire and motivate more valuable research in related fields.



Academic Editor: Paolo Mercorelli

Received: 9 February 2025

Revised: 17 March 2025

Accepted: 20 March 2025

Published: 23 March 2025

Citation: Liu, H.; Li, Y.; Qian, T.; Tang, Y. Recent Progress in Ocean Intelligent Perception and Image Processing and the Impacts of Nonlinear Noise.

Mathematics **2025**, *13*, 1043. <https://doi.org/10.3390/math13071043>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: recent progress; ocean intelligent perception; image processing; nonlinear noise; YOLO

MSC: 68T07

1. Introduction

With artificial intelligence, sensors, and information communication development, ocean environment intelligent perception and target detection [1,2] technology has been comprehensively developed and applied. Intelligent perception of the marine environment relies on intelligent perception devices [3,4], ocean image datasets [5–7], and deep learning algorithms. Intelligent detection devices perceive the marine environment by sending

detection signals and acquiring and processing target images. In order to enhance the environmental perception abilities of intelligent detection devices in complex marine environments, vision-based ocean object detection algorithms have been used as a key technical tool [8]. Zha et al. [9] proposed a deep learning-based computer vision deblurring method to improve the spatial resolution of beamforming images. Sattar et al. [10] presented a vision-processing method for underwater target tracking using various image-processing tools, such as color segmentation, color histograms, and the mean shift. Underwater image processing is a critical issue in ocean intelligent perception [11]. Xu et al. [12] introduced the pre-processing subsystem (GOCI IMPS) of the geostationary ocean color imager (IMage) for Communication, Ocean, and Meteorological Satellites (COMS). They described its functions, development status, and operational concepts. Chen and Liu [13] established a VQA model with an attention mechanism and studied its application in ocean image processing. Furthermore, the advancement of ocean image processing technology is significantly driven by the use of deep learning algorithms. For example, Xue et al. [14] proposed an underwater acoustic recognition and image processing technology based on deep learning, adopting a feature classification method based on one-dimensional convolution to identify target images. Gong et al. [15] proposed a classification method based on deep separable convolutional feature fusion to improve the accuracy of underwater target classification.

However, underwater target detection faces many problems and challenges, such as insufficient sample data, low image resolutions [16], poor signal-to-noise ratios [17], and signal loss in complex underwater environments [18]. Chen et al. [19] pointed out that the interference of ocean turbulence can cause severe information degradation when laser-carrying image information is transmitted in seawater. The specific measures to address this issue mainly include image enhancement and reducing interference. Multi-scale image feature fusion [20–22] provides a practical approach for underwater image enhancement and reconstruction [23]. For example, Gong et al. [24] proposed an underwater image enhancement method based on color feature fusion, which introduces an attention mechanism to design a residual enhancement module to enhance the feature expression ability of underwater images. Chen et al. [25] proposed an underwater crack image enhancement network (UCE CycleGAN) that utilizes multi-feature fusion to reduce blur and enhance texture details. Regarding reducing interference, DL-based wave suppression algorithms are effectively applied in ocean engineering; for example, researchers use DL-based wave suppression algorithms to reconstruct or improve SAR image resolution [26,27]. Jiang et al. [28] designed an improved deep convolutional generative adversarial network (DCGAN) for automatically detecting internal waves in the ocean and reducing their interference. Compared with traditional methods, deep learning algorithms can facilitate higher accuracy and improved ocean intelligent perception image processing performance, especially for radar and sonar detection [29].

Several scholars have reviewed the underwater ocean object detection technologies based on deep learning [30] and image processing with deep learning [31]. Li et al. [32] summarized the key advances in the application of DL algorithms to the visual recognition and detection of aquatic animals, including the datasets, algorithms, and performance. Researchers have also conducted relevant review work [33,34] on ocean intelligent detection and image processing, referring to both acoustic [35,36] and optical image processing [37]. For example, Chai et al. [38] reviewed the research on DL-based sonar image algorithms, including key technologies in denoising, feature extraction, classification, detection, and segmentation. The YOLO series has been widely developed and extensively researched, being a well-known DL-based network model for the processing of ocean intelligent detection images. Wang et al. [39] reviewed the development of YOLOv1 to YOLOv10, focusing on the functional structure improvements. Hussain [40] reviewed the progres-

sive improvements spanning YOLOv1 to YOLOv8, focusing on their key architectural innovations and application performance. Jiao and Abdullah [41] discussed the relevant capabilities of the YOLOv1 to YOLOv7 series algorithms and compared their algorithmic performance. This study comprehensively describes the mathematical structures of the YOLO models, the differences in their network architectures and detection algorithms, and the influence of nonlinear noise from the ocean environment on signal image processing, as well as potential countermeasures. This work provides unique research perspectives and reference value for researchers in related fields. It includes the following areas of focus: ocean intelligent perception devices and image acquisition (Section 2), image recognition and detection models (Section 3), adaptive image processing processes supporting image recognition and detection (Section 4), and coping methods for nonlinear noise interference in image detection (Section 5). Section 6 presents the conclusions and discusses future research directions.

2. Ocean Intelligent Perception Devices and Image Acquisition

Ocean intelligent perception devices, such as autonomous underwater vehicles (AUVs) [42], sonar sensors, LiDAR, and optical cameras, perceive the ocean environment by sending and receiving detection signals. Thus, the acquisition and analysis of signal images have become important aspects of ocean intelligent perception.

2.1. Autonomous Underwater Vehicles with Multiple Sensors

AUVs have been widely applied in the ocean intelligent detection field, enabling the automatic identification and location of targets in diverse marine environments. Integrated with multiple sensors and devices, AUVs can automatically complete various ocean intelligent detection tasks, including multi-target detection, mobile target detection, remote target detection, and real-time detection.

Optical sensors sense the marine environment by sending and receiving optical signals. An optical camera can provide high-resolution real-time images, being suitable for clear-water areas and short-range detection. Wang et al. [43] pointed out that optical sensors can provide more detailed and comprehensive depictions than acoustic sensors in short-range underwater detection. Cameras, which are widely used to detect and track underwater targets [44], have been further enhanced by Rao et al. [45], who proposed a visual detection and tracking system for moving maritime targets using a multi-camera collaborative approach. Yang et al. [46] summarized the use of RGB cameras in intelligent ocean target detection, analyzing their performance in terms of accuracy, speed, and robustness. Combined with multi-sensor systems and advanced machine vision techniques [47], this intelligent marine detection equipment demonstrates powerful detection capabilities.

Acoustic sensors sense the marine environment by sending and receiving acoustic signals. In turbid waters, complex environments, and long-distance detection, the application value of acoustic signal detection in ocean exploration far exceeds that of electromagnetic waves and visual signals [48,49]. Thus, AUVs are usually equipped with sonar sensors [50–52] to detect targets and perceive the complex underwater environment of the ocean. They collect image data and combine them with data from other sensors for a comprehensive analysis, providing more extensive marine environmental information. Acoustic sensors represent a powerful tool and method for ocean intelligent detection, particularly when AUVs are supported by multi-sensor devices and deep learning algorithms [53,54].

Laser sensors perceive the marine environment by sending and receiving laser signals. Usually implemented in AUVs and LiDAR, they stand out due to their ability to quickly acquire high-resolution and accurate detection images during ocean observation and surveys [55,56]. This efficiency makes them more suitable for long-distance and real-time

detection, highlighting their superiority in specific applications, such as underwater laser ranging [57], underwater laser imaging [58], and underwater laser mapping [59].

In summary, AUVs equipped with multi-sensor systems demonstrate remarkable adaptability, as they utilize various signals for ocean target detection, including acoustic, light, electromagnetic, and laser signals. Sonar is the most widely used type of detection signal and is applied in various areas of ocean environment perception. The technical principles and methods of acoustic signal detection are detailed below.

2.2. Sonar Detection Image Acquisition

Sound navigation and ranging (sonar) is the most widely used sensing technology in ocean detection. It harnesses sound waves for a multitude of tasks, from ocean exploration to target recognition, encompassing both active and passive sonar detection, as depicted in Figure 1. Active sonar is a key tool in ocean exploration. It includes forward-looking sonar (FLS), side-scan sonar (SSS), and synthetic aperture sonar (SAS).

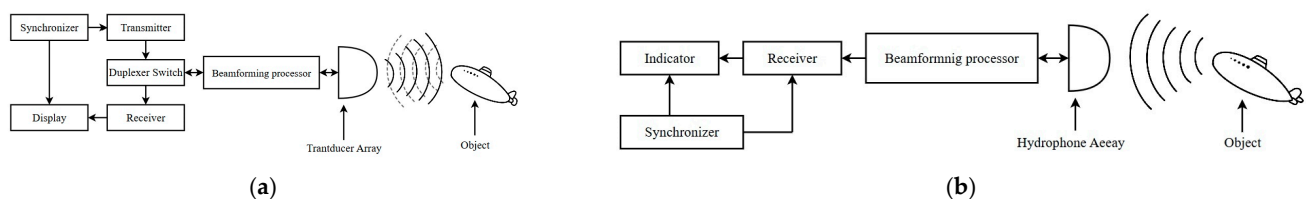


Figure 1. Sonar detection types: (a) active sonar; (b) passive sonar [38].

In ocean intelligent detection, AUVs are equipped with various sensors to perform different tasks; among these, forward-looking sonar (FLS) [60,61], side-scan sonar (SSS) [62], and dual-frequency sonar (DIDSON) [63] are the most common sonar technologies. Forward-looking sonar and side-scan sonar represent basic mapping sensors in underwater detection. Early sonar detection technology enabled the construction of a two-dimensional map for AUVs. For example, the 2D mapping performed via FLS allows AUVs to obtain environmental information in the direction of navigation. FLS is mainly used for forward detection and enables real-time imaging and measurement in front of an AUV. At the same time, high-frequency FLS is useful in detecting and recognizing small underwater objects due to the high-resolution maps that it generates. Side-scan sonar is mainly used for side detection in AUVs; it can provide high-resolution signal images for underwater target recognition and detection. Various sonar sensors provide high-resolution underwater images to assist AUVs in detecting underwater terrain and landforms. Combined with FLS and SSS, AUVs can obtain 3D mapping information from sonar detection signals [64]. FLS and SSS can provide real-time detection information for AUVs in different directions in complex underwater environments. Taking SSS as an example, the real-time acoustic detection process of sonar sensors is shown in Figure 2.

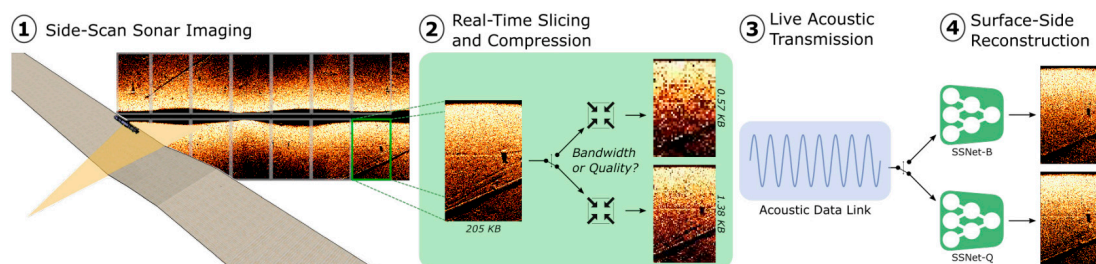


Figure 2. The real-time acoustic detection process of side-scan sonar [65].  represents the image compression process.

Dual-frequency sonar is a sonar technology that uses two different frequencies (1.8 and 1.1 MHz) of sound waves for detection and identification. It was first developed by the Applied Physics Laboratory of the University of Washington for the United States Navy [66]. By using two different frequencies of sound waves simultaneously or alternately, dual-frequency sonar can obtain more information, providing real-time underwater images with near-video quality [67]. Combined with a coaxially mounted low-light video camera (Figure 3a), DIDSON can obtain relatively high-quality detection images by using a more extensive sampling window (Figure 3b); this has been widely applied in marine fish detection [68–70].

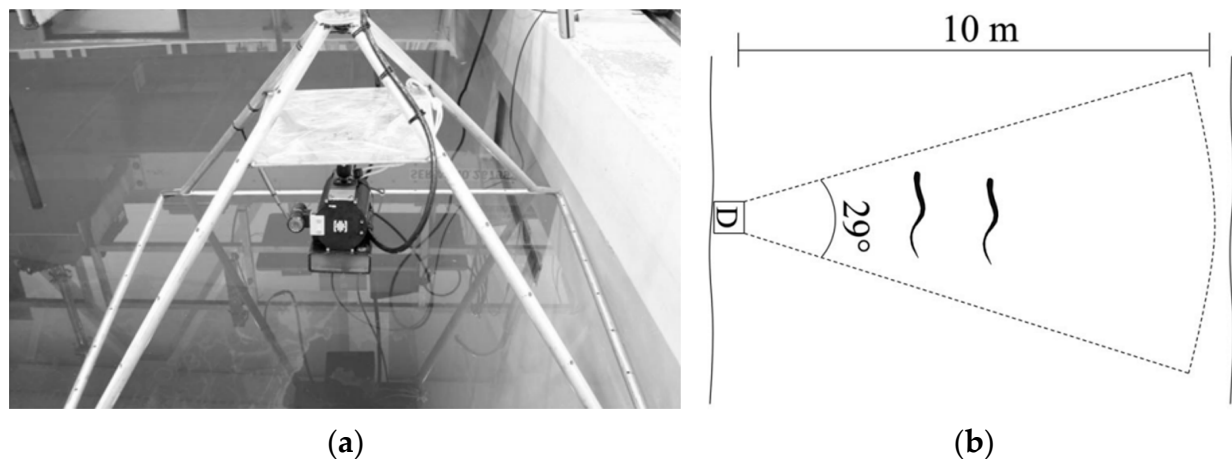


Figure 3. Working process of DIDSON: (a) low-light video camera [71]; (b) larger sampling window of dual recognition sonar [72].

2.3. Resource Limitations of Ocean Intelligent Perception Devices

Underwater sensors have significant resource constraints from energy storage and communication methods when detecting underwater targets [73]. For example, the wireless communication performance of underwater wireless sensor network (UWSN) nodes is poor in underwater environments [74]. The complex and ever-changing marine environment makes it difficult for underwater fiber-optic acoustic sensors to achieve high-precision identification, positioning, and tracking [75]. The insufficient real-time performance and refresh rate of underwater acoustic positioning and navigation systems further exacerbate the situation [76]. The fusion problem of multi-source information data makes it difficult for autonomous underwater vehicles (AUVs) to perceive the external environment accurately [77]. Insufficient capacity of edge devices for underwater data collection entities leads to difficulties in processing complex data [78]. Lastly, the nonlinear noise interference caused by underwater radiated noise (URN) [79] significantly affects the detection accuracy of underwater sensing equipment.

The above resource constraints pose significant challenges to high-precision detection, real-time detection, and multi-modal fusion detection of ocean intelligent perception, requiring more advanced intelligent image processing algorithms, multi-sensor fusion methods [80], denoising methods, and lightweight AI [81,82] models to address these challenges.

3. Ocean Image Recognition and Detection Models

DL-based network models are used in ocean intelligent perception and detection to analyze and process the signal images received by optical, acoustic, or laser sensors. Deep learning network models have played an important role in ocean detection signal processing, encompassing tasks such as image denoising [83], feature extraction [84], classification [85], recognition [86], and detection [87] across sonar images [88,89], optical

images [90], and laser images [91]. Chai et al. [38] reviewed the applications of various deep learning networks in image processing, providing a thorough analysis of the relevant types involved.

From the perspective of technological development, underwater target feature extraction and recognition technology has progressed from traditional time–frequency-domain methods to deep learning algorithms and end-to-end network models. The former type of method relies on a manually designed feature extraction process, which can easily lead to information loss and poor environmental adaptability. The latter two types of methods both rely on deep learning network models. Yang et al. [92] presented a significant advancement in the field, namely the end-to-end underwater acoustic target recognition model (1DCTN), which could achieve a relatively high recognition rate. Multi-target detection typically relies on deep learning models and devices such as cameras and uncrewed boats to detect ocean targets. Underwater object detection technology is evolving rapidly, as shown in Figure 4, and the 1DCTN model is a key player in this transformation, providing crucial support for autonomous devices in ocean environmental perception and detection [93].

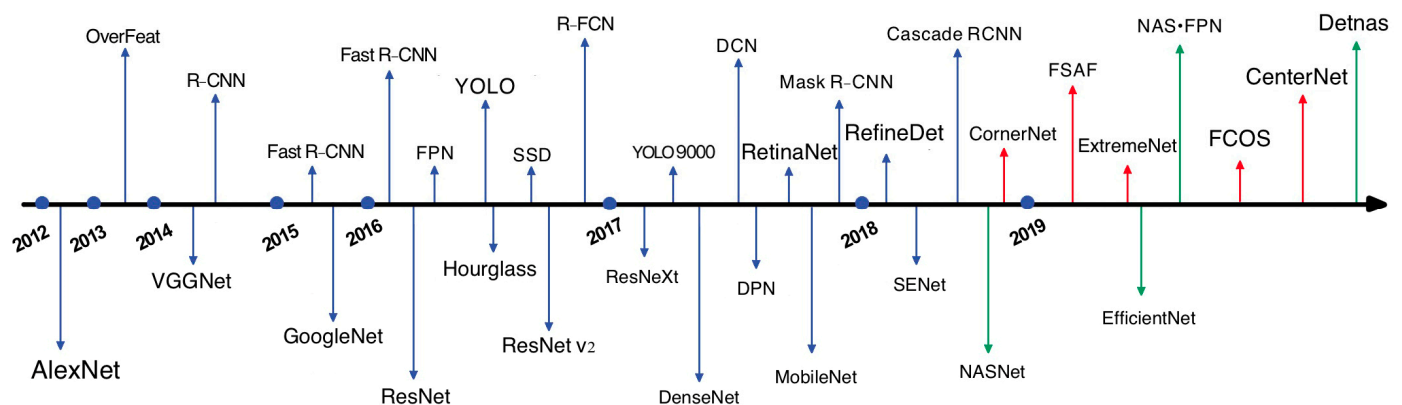


Figure 4. The object detection milestones [41].

3.1. Deep Convolutional Neural Networks

Deep convolutional neural networks (CNNs) [94–97] are the most well-known network models used for image feature extraction and recognition. These are typically used to extract image features to capture local and global visual information in ocean target detection. CNNs are usually applied in marine acoustic detection and image analysis [98–100], such as signal image recognition and classification for sonar or synthetic aperture radar (SAR) [101]. These models can learn features from a large amount of annotated data, significantly improving the accuracy and robustness of detection. For example, Ashok and Latha studied the use of convolutional neural networks to extract underwater acoustic signal image features by constructing an innovative DNN called the Audio Perspective Region-Based Convolutional Neural Network (APRCNN) [102]. Wang and Li [103] discussed the application of a CNN and its variants in the processing and analysis of ocean remote sensing data. However, conventional object detection algorithms rely on feature-based methods such as HOG, SIFT, and SURF and are unsuitable for dynamic and real-time target detection in complex scenes. When basic CNNs are applied to ocean image recognition using small samples, challenges arise due to the marine environment's complexity, background interference, the diversity and variability of marine targets, uneven datasets, and the requirements for real-time detection. Despite these challenges, the potential of CNNs to improve the accuracy and robustness in complex underwater environments is evident. Nevertheless, the DCNN is a complex model with many convolutional and pooling layers. It will require large amounts of inference time and computation when detecting ocean

targets. Thus, more lightweight and accurate models are needed for real-time detection in complex underwater environments.

3.2. Two-Stage Detection Network Model

Due to their high detection precision, efficiency, and flexibility, two-stage detection network models have been widely developed and applied in ocean intelligent detection. The specific work process includes the following steps. The first is to use the selective search algorithm [104] to generate a series of candidate region proposals from the input image, which may contain the target object. The second step includes extracting features from each candidate region using pre-trained convolutional neural networks such as AlexNet [105] and VGG [106]. These features can be used to describe the content within the region. In the last step, the extracted features are input into a support vector machine (SVM) [107] for classification, seeking to determine whether the candidate region contains the target object. At the same time, a regression model is used to perform bounding box regression on the candidate regions so as to improve the accuracy in identifying the target position. Bounding box regression is a technique that adjusts the bounding box coordinates to fit the object better. The R-CNN series consists of several well-known two-stage object detection algorithms [108,109], including Faster R-CNN, Mask R-CNN, Cascade R-CNN, and other improved models. Ocean target detection based on R-CNN can be divided into the following steps: selective search, feature extraction, classification, and regression. However, the main limitation of R-CNN in ocean target detection is that its computational efficiency is relatively low, because each region proposal requires separate feature extraction through the CNN. In addition, the computational cost of the selective search algorithm is significant.

Faster R-CNN has been improved based on R-CNN through the introduction of the Region Proposal Network (RPN). The RPN generates region proposals directly through the CNN, avoiding the selective search and significantly improving the speed. Faster R-CNN still uses the CNN for feature extraction and classification but integrates the region proposal generation process into the network. Therefore, Faster R-CNN allows significant improvements in both speed and accuracy, making it more suitable for multi-target and real-time detection. Moreover, it can quickly adapt to different marine environments and target types, such as SAR image detection [110], underwater target detection [111], and marine debris and biological exploration [112–114].

Mask R-CNN [115,116] is a versatile extension of Faster R-CNN, significantly enhancing its capabilities by adding an instance segmentation functionality. Pixel-level segmentation is enabled by introducing a branch based on Faster R-CNN to produce binary masks for each target. Due to its excellent image segmentation abilities, Mask R-CNN has been widely adopted in various marine target detection tasks, such as SAR image detection [117,118], marine debris exploration [119], and marine biological detection [120,121]. For instance, Mana and Sasipraba [122] demonstrated the power of Mask R-CNN by developing a deep learning-based classification technique using a mask-based region convolutional neural network for fish detection and classification.

Furthermore, the development of improved R-CNN models for object detection has been undertaken as a collaborative effort, with practical applications in various fields. For example, Bi et al. [123] contributed the Information-Enhanced Mask R-CNN (IEMask R-CNN), which uses the FPN structure; this enhances the valuable channel information and global information of feature maps. This model is particularly useful for segmentation tasks in computer vision. Lee et al. [124] proposed an ME R-CNN with an expert assignment network (EAN), which captures various areas of interest in computer vision. Liu et al. [125] explored an SE-Mask R-CNN by introducing a squeeze-and-excitation block

into the ResNet-50 backbone, improving the speed and accuracy of fruit detection and segmentation. These advancements in R-CNN models have been further extended to enhance the performance of underwater target detection models, such as Boosting R-CNN [108] for two-stage underwater exploration, Mask-Box Scoring R-CNN [126] for sonar image instance segmentation, and Coordinate-Aware Mask R-CNN [127] for fish detection in commercial trawl nets. Figure 5 illustrates the development of R-CNN and its partially improved models in ocean target detection.

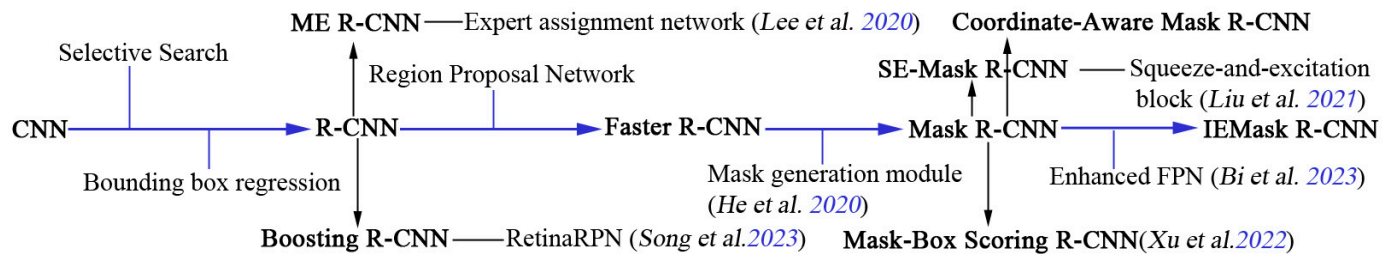


Figure 5. The development of R-CNN and its partially improved models in ocean target detection [108,115,123–126].

3.3. Single-Stage Detection Network Model

The network models mentioned above enable the accurate identification of targets, but the computational complexity is too high for embedded systems, reducing the inference speed. Even for the fastest high-precision detector, Faster RCNN can only run at a speed of 7 frames per second, significantly reducing the update efficiency in real-time detection. Therefore, an end-to-end object detection method using single-stage detection algorithms has been developed using the SSD and YOLO series, which are well-known network models.

Single-Shot Multi-Box Detector (SSD) [128,129] is a single-stage detection algorithm that differs from traditional two-stage detection algorithms such as Faster R-CNN. SSD does not require the generation of candidate regions; it only needs an input image and ground truth boxes for each object. It performs well in applications requiring high real-time performance, striking an appropriate balance between speed and accuracy. SSD excels in real-time detection [130] and multi-target detection [131]. SSD exhibits a faster inference speed and is superior in tasks such as video surveillance, autonomous driving, and ocean exploration. Moreover, SSD uses multi-scale feature maps to detect targets, enabling it to capture targets of varying sizes by predicting feature maps of different levels, maintaining relatively high accuracy.

YOLO is a well-known computer vision language model for real-time object detection that was developed by Joseph Redmon and his team [132]. With continuous upgrading and improvement, YOLO has evolved into a fast, real-time, multi-target detection algorithm. Its high precision makes it a reliable choice for real-time detection, such as in autonomous driving and robot vision. Due to its robust real-time detection and image analysis capabilities, YOLO has been widely used in marine object detection [133,134]. The YOLO model differs significantly from other network models in terms of network structure and inference methods. It has a higher inference speed but a lower detection accuracy. Table 1 shows a comparative study between YOLO with the two-stage detection network of Faster R-CNN and the single-stage detection network of SSD. However, by improving its mathematical structure, the YOLO model is constantly breaking through and upgrading detection accuracy, developing multiple advanced versions.

Table 1. Comparative study between YOLO and other common detection models.

Network Model	Inference Method	Advantage	Limitation	Adaptation Scenarios
Faster R-CNN	Generate candidate regions and then classify and regress them	High precision and excellent performance in complex scenes	Slow speed, difficult to meet real-time requirements	Scenarios that require high precision but not strict speed requirements
SSD	Directly output target's category and location, using multi-scale feature maps	Fast speed, suitable for real-time applications	Poor performance in detecting small targets	Real-time detection with moderate precision requirements
YOLO	Divide the image into grids, and each grid predicts multiple bounding boxes and category probabilities	Extremely fast, suitable for high real-time requirements	Relatively low accuracy, poor performance in detection of small and dense targets	Scenarios with extremely high speed requirements

3.4. Mathematical Structures of YOLO Model

In ocean image detection tasks, the superiority of the YOLO model over other object detection models, such as Faster R-CNN, SSD, and RetinaNet [135], is mainly reflected in the following aspects. First is the real-time detection ability. Due to the adoption of a single-stage detection architecture, the inference speed has been dramatically improved, making it suitable for ocean real-time monitoring. Second is the multi-scale prediction and small object detection capability. Due to the use of multi-scale feature fusion and anchor prediction mechanism, its detection ability of small targets is greatly enhanced, which is suitable for detecting marine plankton and floating garbage. Third is the lightweight network architecture design. Some YOLO variants, such as YOLOv5s and YOLOv8n, compress models through channel pruning, quantization, and other techniques, with parameter sizes of only 1–7 M, and can be deployed on edge devices such as buoys and underwater robots. Thus, the mathematical structure of the YOLO model achieves efficient detection through a lightweight backbone network, multi-scale feature fusion and grid prediction, and composite loss function.

1. Lightweight backbone network

YOLO uses convolutional neural networks (CNN) as its essential backbone. The earliest YOLO model (YOLOv1) includes 24 convolutional layers followed by two fully connected layers, as shown in Figure 6.

Regardless of the network structure, the core algorithm is the convolution operation, yielding

$$O(x, y) = f(\sum_i \sum_j I(x + i, y + j) \cdot W(i, j) + b) \quad (1)$$

where $I(x, y)$ and $O(x, y)$ represent the input and output pixels, respectively. $W(i, j)$ represents the convolutional kernel weights with i and j steps. b is the offset value, and f is the activation function, e.g., Leaky ReLU, yielding

$$f(x) = \max(x, 0.1x) \quad (2)$$

Subsequent YOLO versions have developed more complex network structures as their backbone network based on this foundation. Their network backbone reduces computational complexity and improves inference speed by introducing Darknet and CSPDarknet to achieve cross-stage partial networks. They also integrate multi-scale contextual information by introducing SPP/ASPP modules to enhance robustness against complex ocean backgrounds.

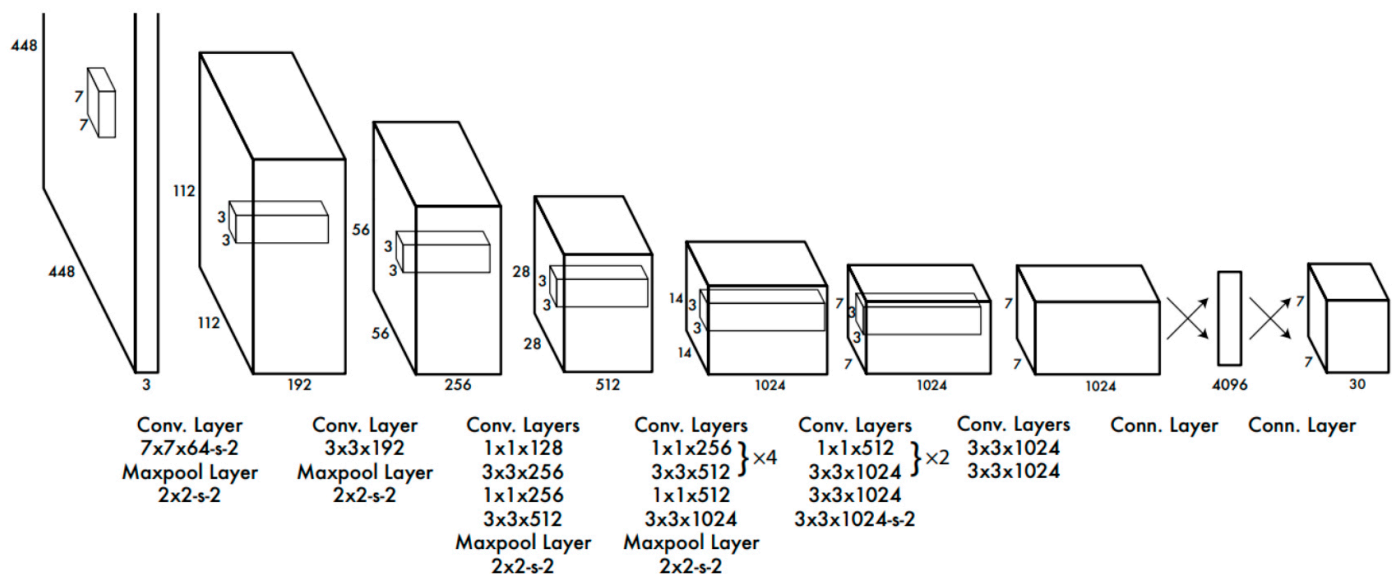


Figure 6. The YOLO architecture [132]. The “-s-” means the stride.

2. Multi-scale feature fusion and grid prediction

• Multi-scale prediction;

Multi-scale prediction is a critical method for YOLO models that enables the model to detect large and small targets simultaneously. This method improves the model’s adaptability to targets of different sizes and detection accuracy through multi-scale feature fusion. The implementation of multi-scale prediction involves the introduction of anchor boxes, multi-scale training, and other network structures. YOLOv3 and its corresponding versions achieve multi-scale prediction by integrating feature pyramid networks (FPNs) [136] or path aggregation networks (PANs) [137]. These networks play a crucial role in the field by improving feature aggregation, enabling context information in images to be captured by aggregating feature maps from different layers. Dong et al. [8] analyzed the structures of FPN and PAN, as shown in Figure 7. The FPN introduces a top-down pathway for the fusion of multi-scale features on top of the SSD, while the PAN adds a bottom-up pathway on top of the FPN.

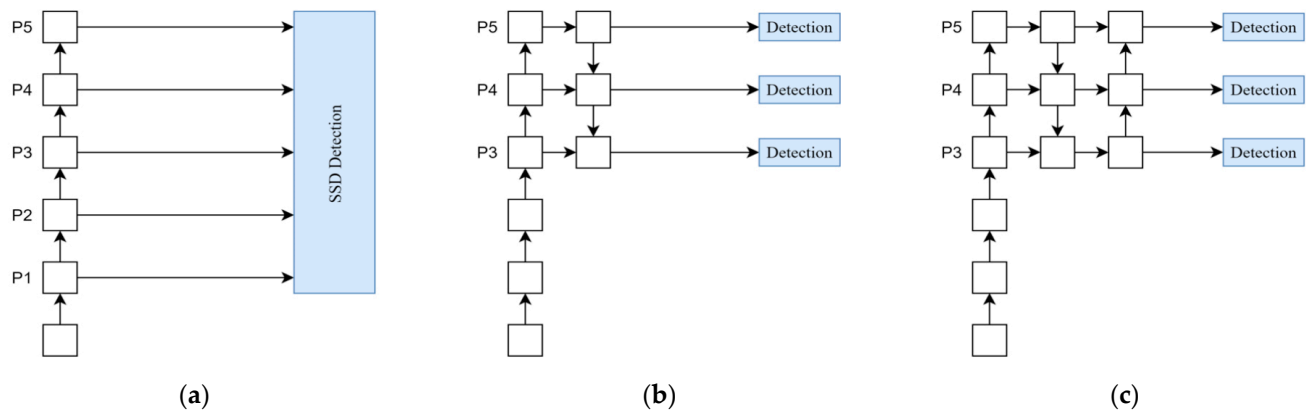


Figure 7. Different types of feature fusion networks: (a) SSD; (b) FPN; (c) PAN.

Multi-scale feature fusion, the basis of multi-scale prediction, integrates features from different levels, empowering the model to confidently handle targets of various scales. Signal image processing network models usually incorporate multi-scale feature fusion structures after the convolutional layer, pooling feature maps of different scales to extract multi-scale feature information. This strategic design effectively overcomes the limitations of traditional convolutional neural networks in processing images of different sizes, providing a reassuring level of adaptability.

- Grid prediction;

YOLO models use the anchor mechanism to solve the ambiguity problem of dense targets in marine scenes. This mechanism dynamically allocates positive samples by jointly measuring the prediction box and truth value, ensuring dynamic prediction and adaptive labeling. The anchor mechanism further enhances YOLO's capabilities by dividing the image into grids, allowing each grid cell to predict multiple bounding boxes, each corresponding to an anchor box. It uses pre-defined anchor boxes to predict the locations and categories of targets. Each anchor box has a specific aspect ratio and scale, and the algorithm performs bounding box regression. This involves predicting the bounding box of the target through regression analysis, where the coordinates of candidate regions are adjusted to surround the target. The process also includes target classification and label assignment. Taking YOLOv3 as an example, the center point coordinates of the predicted bounding box can be represented as

$$\hat{x}_i = \sigma(x_i) + x \quad (3)$$

$$\hat{y}_i = \sigma(y_i) + y \quad (4)$$

where (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ denote the center point coordinates of the true and predicted bounding box, respectively. (x, y) is the coordinate in the upper-left corner of the grid. σ represents the sigmoid function. The predicted bounding box size can be expressed as

$$b_w = w_i e^{t_w} \quad (5)$$

$$b_h = h_i e^{t_h} \quad (6)$$

where (w_i, h_i) represents the preset anchor size. (t_x, t_y) represents the original pixel coordinates output by the model, and (t_w, t_h) represents the true bounding box size. A detailed explanation of this process is shown in Figure 8.

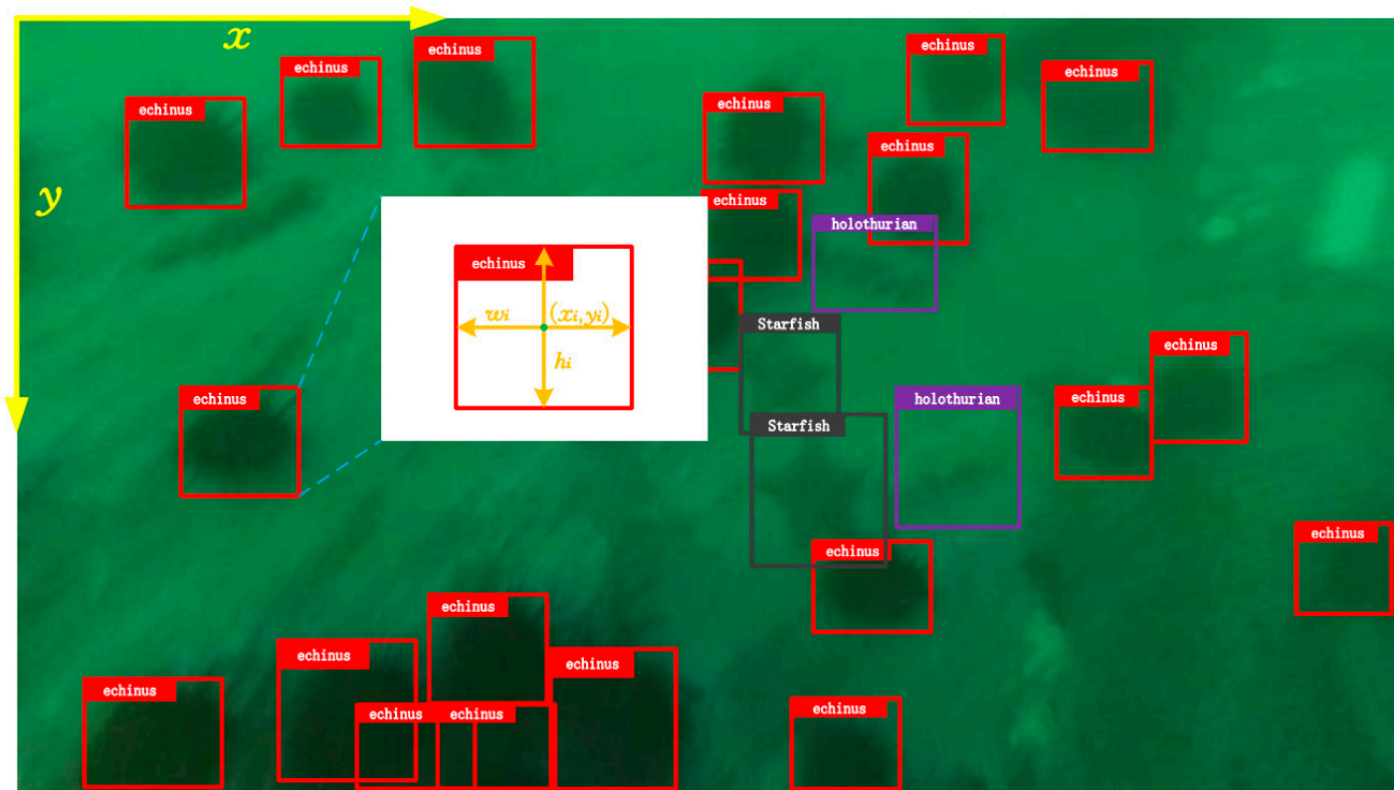


Figure 8. Process of predicting bounding box and category label for object of interest, where (x_i, y_i) denotes the coordinates of the center point of the bounding box predicted by the model [32].

3. Composite loss function

The design of the loss function is crucial for the YOLO model's object detection performance. The loss function usually consists of positioning loss, classification loss, and object confidence loss.

- Positioning loss;

The positioning loss calculates the difference between the predicted bounding box and the truth bounding box, guiding the model to adjust the center coordinates (x, y) and dimensions (w, h) of the bounding box to make it closer to the truth target position. The mean squared error (MSE) [138] and CIoU loss function are usually used to calculate the positioning loss. For example, it uses the MSE loss function to calculate the positioning loss from YOLOv1 to YOLOv3, yielding

$$L_{loc-MSE} = \lambda_{coord} \sum_{i=1}^{S^2} \sum_{j=1}^N 1_{ij}^{noobj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \quad (7)$$

where (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ denote the center point coordinates of the i -th true and predicted bounding boxes, as mentioned in Equations (3) and (4). S^2 represents the number of grids that the image is divided into, $S \times S$; 1_{ij}^{noobj} is the exponential function used to screen negative samples. λ_{coord} is a weight coefficient used to amplify the importance of the localization loss.

It uses the CIoU loss function to calculate the positioning loss from YOLOv4 to YOLOv8, yielding

$$L_{loc-CIoU} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (8)$$

where IoU is the Intersection over Union of predicted and truth bounding box; $\rho^2(b, b^{gt})$ represents the Euclidean distance between the predicted box center point b and the truth

box center point b^{gt} ; c is the diagonal length of the smallest bounding rectangle between the predicted box and the truth box; α is the weight coefficient used to balance the penalty term of aspect ratio; and v is the penalty term of aspect ratio, which yields

$$v = \frac{4}{\pi^2} (\arctan(\frac{w^{gt}}{h^{gt}}) + \arctan(\frac{w}{h})) \quad (9)$$

where (w, h) and (w^{gt}, h^{gt}) represent the width and height of the predicted box and truth box, respectively.

- Classification loss;

The classification loss measures the difference between the class probability distribution predicted by the model and the truth class labels. It plays a critical role in guiding the model in adjusting the category prediction results, ensuring that each target's category prediction is as accurate as possible. The binary cross-entropy (BCE) loss function [139] is the method most commonly used to calculate the classification loss, which yields

$$L_{class-BCE} = -\frac{1}{N} \sum_{i=1}^N [p_i \log \hat{p}_i + (1 - p_i) \log (1 - \hat{p}_i)] \quad (10)$$

where p_i represents the true label of the i -th box with a value of 0 or 1. $\hat{p}_i$ represents the probability value predicted by the model of the i -th box, ranging from 0 to 1. N is the sample number.

- Confidence loss;

The confidence loss is used to evaluate the model's prediction accuracy, i.e., whether the predicted bounding box contains the target. In the YOLO series, the MSE or BCE loss function can be used to calculate the confidence loss. The confidence loss calculation by using the MSE loss function yields

$$L_{Conf-MSE} = -\sum_{i=1}^{S^2} \sum_{j=1}^N 1_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=1}^{S^2} \sum_{j=1}^B 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 \quad (11)$$

where λ_{noobj} is a weight coefficient used to balance the confidence loss of positive and negative samples. C_i represents the true confidence score of the i -th box, and $\hat{C}_i$ represents the predicted confidence score of the i -th box, yielding

$$C_i = \text{Pr}(\text{Object}) \times \text{IoU}_{\text{pred}}^{\text{truth}} \quad (12)$$

where $\text{Pr}(\text{Object})$ denotes the probability that the bounding box contains the target. Both C_i , $\hat{C}_i$ and $\text{Pr}(\text{Object})$ can take a value of 0 or 1. $\text{IoU}_{\text{pred}}^{\text{truth}}$ is the intersection over union ratio of the true and predicted bounding boxes.

The confidence loss calculation by using the BCE loss function yields

$$L_{Conf-BCE} = -\sum_{i=1}^{S^2} [C_i \log \hat{C}_i + (1 - C_i) \log (1 - \hat{C}_i)] \quad (13)$$

3.5. Development and Comparative Study of YOLO Models

The mathematical structure of the YOLO model plays a crucial role in determining its detection accuracy and inference speed. This has led to the development of multiple versions of the YOLO network model, of which the official versions range from YOLOv1 to YOLOv8, which have been maturely applied in diverse fields, including ocean intelligent detection. The latest versions, YOLOv9, YOLOv10, and YOLOv11, are unofficial (community-improved) versions that adopt the backbone network architecture of YOLOv8 and have further optimization. YOLOv9 introduces reversible networks to achieve a

lightweight design [140]. In YOLOv10 and YOLOv11, the models' lightweight nature and real-time detection capabilities are optimized [141].

The different YOLO versions vary in their mathematical structure, which is evident in their network architecture and detection algorithms, as shown in Table 2.

Table 2. Mathematical structure of YOLO official versions from YOLOv1 to YOLOv8.

Model Version	References	Network Architecture	Multi-Scale Prediction	Activation Function	Loss Function
YOLOv1	Redmon et al. [132]	FCN	Anchor-Free	Leaky ReLU	MSE for location, classification, and confidence
YOLOv2	Redmon et al. [142] Shafiee et al. [143] Panchal et al. [144]	Darknet-19	Anchor-box	Leaky ReLU	MSE for location; BCE for classification
YOLOv3	Redmon et al. [145] Gunawan et al. [146] Gaussian YOLOv3 [147]	Darknet-53	Anchor-box+ FPN	Leaky ReLU	MSE for location; BCE for classification and confidence
YOLOv4	Bochkovskiy et al. [148] Liu et al. [149] Scaled YOLOv4 [150]	CSPDarknet-53	Anchor-box+ PAN	Mish	CIoU for location; BCE for classification and confidence
YOLOv5	Ge et al. [151] Zendehdel et al. [152]	CSPDarknet-53	Anchor-box+ PAN + FPN	SiLU	CIoU for location; Focal loss for classification and confidence
YOLOv6	Gupta et al. [153] Li et al. [154] YOLOv6-ESG [155]	CSPDarknet-53	Anchor-Free+ PAN + FPN	ReLU	CIoU/GIoU for location; Focal loss for classification; BCE for confidence
YOLOv7	Cai et al. [156] Jiang et al. [157] Patel et al. [158]	Extended CSPDarknet-53	Anchor-box+ PAN + FPN; Introducing more cross-scale connections	SiLU	CIoU for location; BCE for classification and confidence
YOLOv8	Song et al. [159] Zhang et al. [160] Qu et al. [161]	Extended CSPDarknet-53 (optimized version)	Anchor-Free+ PAN + FPN; Further optimization of cross-scale connectivity and feature transfer	SiLU	CIoU for location; BCE for classification and confidence; DFL for classification optimization

Furthermore, due to the differences in mathematical structures between different versions, their detection performance also varies. This is specifically manifested in the adaptability of the network architectures of different YOLO versions to different tasks in ocean target detection, localization, classification, and image recognition, as shown in Table 3.

Table 3. Performance of YOLO official versions from YOLOv1 to YOLOv8.

Model Version	Detection Accuracy (mAP)	Inference Speed (FPS)	Resource Consumption (Parameter Quantity/Model Size)	Detection Ability Performance	Ocean Perception Scene Adaptability
YOLOv1	63.4% (VOC 2007 test)	45	About 7.5 M/25 MB	Moderate positioning ability, basic classification (only 20 categories), high real-time performance, but low accuracy	Simple static target detection (large sunken ship)
YOLOv2	76.8% (VOC 2007 test)	67	About 50 M/67 MB	High positioning ability, supporting more categories, balance between speed and accuracy	Medium-scale target detection (submersibles, buoys)
YOLOv3	55.3% (COCO AP50-95)	30–50	About 62 M/236 MB	High positioning ability, enhanced classification ability (80 categories), strong robustness in complex scenes	Multi-scale targets (fish schools, floating objects)
YOLOv4	65.7% (COCO AP50-95)	50–80	About 64 M/244 MB	Extremely high positioning ability, multi category classification optimization, suitable for high-resolution ocean images	High-precision underwater terrain mapping
YOLOv5	68.9% (COCO AP50-95)	100–140	About 7.0 M/27 MB (YOLOv5s)	Extremely high positioning ability, supporting custom categories, lightweight model suitable for edge devices	Underwater fuzzy targets and real-time ocean monitoring
YOLOv6	69.5% (COCO AP50-95)	300–500	About 12 M/38 MB (YOLOv6n)	High positioning ability, improved classification efficiency, suitable for low computing power scenarios	Real-time detection optimization and dynamic ocean target tracking
YOLOv7	71.3% (COCO AP50-95)	160–200	About 37 M/71 MB	Extremely high positioning ability, fine-grained classification capability, excellent in handling occluded targets	High-density object detection (fish schools, coral reefs)
YOLOv8	72.5% (COCO AP50-95)	180–220	About 3.2 M/6.2 MB (YOLOv8n)	High positioning ability, classification-independent optimization, balanced speed, and accuracy	Strong universality and adaptability to the diversity of marine targets

For the unofficial versions from YOLOv9 to YOLOv11, the detection performance needs to be verified using specific scenarios. Sharma and Kumar [162] compared the detection performance and efficiency of YOLOv8 to YOLOv11 in practical applications. Moreover, they offer powerful tools for practical applications. Yuan et al. [163] explored the application of YOLOv9 in signal processing for underwater detection scanning sonar; they conducted ablation experiments that confirmed its superiority over other models such as Faster RCNN and YOLOv5. Similarly, Tu et al. [164] developed a fully optimized YOLOv10 model for diseased fish detection, demonstrating the potential of these models in real-world scenarios.

4. Adaptive Image Processing Processes Supporting Ocean Image Recognition and Detection

In ocean image recognition and detection, specific adaptive image processing processes provide critical support for ocean image recognition and detection, which forms a complete process from data preprocessing to image analysis. These image-processing processes are realized by introducing related DL-based network structures.

4.1. Adaptive Image Annotation

As an important part of the data training and preparation phase, underwater image data annotation refers to manually or semi-automatically labeling and annotating ocean environment images or video data obtained through underwater detection equipment. This process aims to provide structured and labeled training data for deep learning models to support ocean image recognition and target detection tasks. It annotates target object images to classify them or calibrates the position (bounding box), contour (semantic segmentation), or pixel-level annotation (instance segmentation) of the target for target tracking and recognition.

However, traditional manual image annotation has drawbacks, such as low efficiency in time- and labor-consuming tasks, inconsistent annotation standards, environmental interference in annotation quality, and difficulty annotating small and dense targets. Intelligent technology methods and tools have been developed and widely used to cope with these issues in underwater image analysis and annotation, such as marine image annotation software (MIAS) [165]. Schoening et al. [166] considered past observations in manual ocean image annotation (MIA), analyzed the performance of human expert annotation, and compared it with computational methods using a statistical framework. Giordano et al. [167] explored a diversity-based clustering technique for image annotation. They achieved remarkable testing performance on a large fish image dataset, ensuring the diversity of annotation items while maintaining accuracy. In adaptive image annotation, pre-trained network models such as Mask R-CNN [168–170] are usually used to guide underwater image annotation automatically, significantly improving the efficiency. For example, Wu et al. [171] proposed an image localization perception intelligent annotation algorithm LACP AL based on the Faster R-CNN active learning framework to improve the recognition accuracy of underwater image detection models.

4.2. Adaptive Image Feature Enhancement

Underwater target detection and image recognition encounter many challenges, such as signal attenuation caused by the seawater environment. These factors result in low contrast, poor clarity, and color degradation in underwater images. Image feature enhancement technology helps to enhance image features and mitigates the impact of environmental interference, improving the recognition and detection capabilities of relevant network models. Deep learning network models play an important role in adaptive enhancement of

ocean images. Lisani et al. [172] evaluated seven state-of-the-art techniques for underwater image enhancement and ranked their practicality in the annotation process: InfoLoss, BAL, Fusion, UCM, ARC, UDCP, and MSR. It often enhances image features by introducing attention modules in image detection models. Taking the well-known YOLOv7 as an example, scholars have introduced attention mechanisms to enhance its image detection performance, as shown in Table 4.

Table 4. Attention mechanisms introduced in YOLO model and their and enhancement effects.

References	Attention Mechanism	Enhancement Effect
Li et al. [173] Ren et al. [174] Chen et al. [175]	Convolutional block attention module (CBAM)	Enhanced channel and spatial dimension features
Yang et al. [176]	Cross-modal Transformer attention	
Liu et al. [177] Qin et al. [178]	Global attention mechanism	Multi-scale feature fusion and enhancement
Yu et al. [179]	3D attention mechanism	Improved anti-interference abilities in underwater recognition
Fu et al. [180] Gao et al. [181]	Coordinate attention (CA)	Enhanced spatial information and prevention of feature loss
Zhao et al. [182]	Simple parameter-free attention	Combination of channel domain and spatial domain
Yi et al. [183]	SENet attention mechanism	Enhanced feature expression and small-target feature information extraction capabilities
Lu et al. [184]	SimAM attention mechanism	
Zhao et al. [185] Chen et al. [186]	Channel attention module (SE)	Enhanced adaptive feature extraction capabilities
Ou et al. [187]	LSKA attention mechanism	Enhanced multi-scale feature extraction capabilities

4.3. Adaptive Image Segmentation

As the core task of ocean image recognition and detection, image segmentation provides structured data support for subsequent analysis and decision-making in ocean image detection by separating targets and backgrounds. This technology divides the image into several subregions with similar attributes by leveraging the features of pixels or regions, such as color, texture, shape, and brightness. It achieves adaptive segmentation of ocean images through deep learning network models, including instance segmentation and semantic segmentation.

- Instance segmentation

Instance segmentation means pixel-level segmentation, which is achieved through joint learning of target localization, classification, and pixel-level segmentation. It provides high-precision data for ocean image detection by accurately distinguishing and labeling each independent target instance in the image. The best-known adaptive instance segmentation models include Mask R-CNN [188], YOLACT [189], and the SOLO series [190]. The primary

purpose of semantic segmentation is to classify different regions of an image by assigning a category label to each pixel.

- Semantic segmentation

Semantic segmentation is a technique that classifies each pixel in an image into a specific category, such as “coral”, “seaweed”, “sand”, “fish”, and “plastic waste”. Unlike instance segmentation, semantic segmentation does not distinguish individual instances of the same target type but focuses on semantic parsing of the global scene. The best-known adaptive semantic segmentation models include FCN [191], path aggregation network (PAN) [192], PSPNet [193], and SegNet [194], which are not simply tools but technological innovations; they enable us to effectively segment targets of interest from signal detection images and process images with complex backgrounds. For example, Yu et al. [195] proposed a multi-attention PAN with a pyramid network for the semantic segmentation of marine images, as shown in Figure 9. In this model, the intelligent perception network model combines the FPN backbone and bottom-up path enhancement module to significantly enhance the network model’s image recognition and semantic segmentation capabilities.

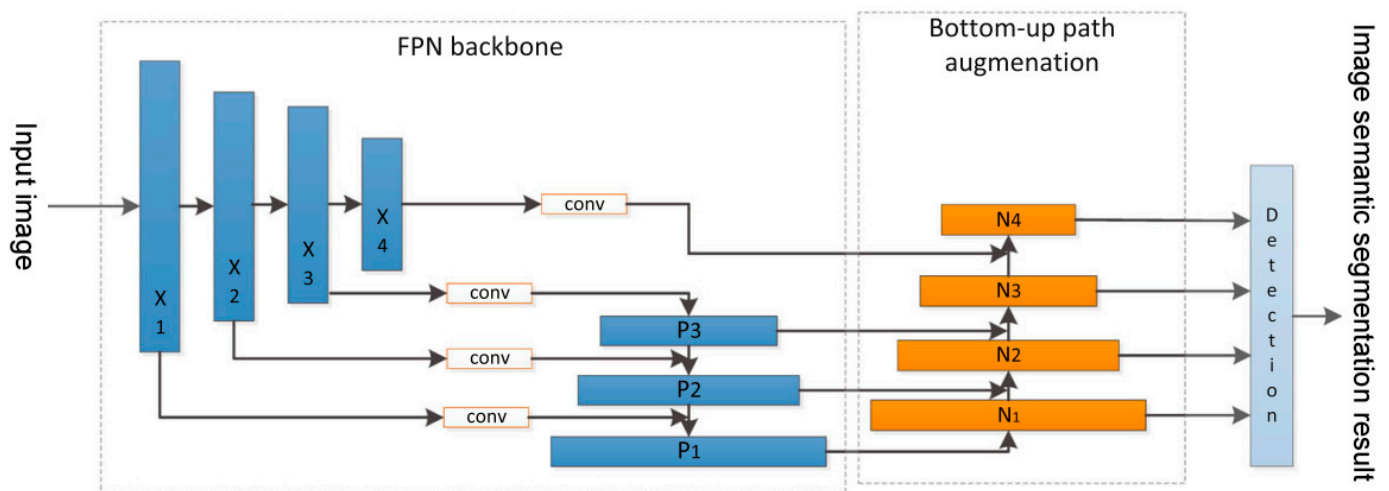


Figure 9. Multi-attention path aggregation network with pyramid network.

- Scene applicability of adaptive image segmentation models

Different network structures are suitable for different semantic segmentation tasks. For example, UNet [196–198] is suitable for small sample image semantic segmentation. Hasimoto-Beltran et al. [199] proposed a multi-channel DNN model based on UNet and ResNet [200] to realize the pixel-level segmentation of marine oil pollution detection images, while the DeepLab series [201] are suitable for high-precision image semantic segmentation. Among them, DeepLabv3+ is the most popular advanced network model, being widely used in ocean intelligent detection [202–204]. However, high-precision and complex semantic segmentation tasks consume a significant amount of computational resources. To address these challenges, a lightweight end-to-end semantic segmentation network combined with attention and feature fusion modules has been developed and applied [205]. For example, Wang et al. [206] developed a lightweight multi-level attention adaptive feature fusion segmentation network (MA2Net) that integrates the proposed lightweight attention network (LAN) with multi-scale feature pyramid (MASPP) and adaptive feature fusion (AFF) models. This model enhances the ability of neural networks to segment target images in an end-to-end manner, without relying on complex DCNN models, and it realizes fast inference and high-precision semantic segmentation.

5. Impact of Nonlinear Noise on Ocean Signal Detection and Countermeasures

5.1. Impact of Nonlinear Noise on Ocean Signal Detection

The generation of ocean nonlinear noise [207] involves multiple complex physical processes, mainly derived from nonlinear effects in fluid dynamics, sound wave propagation, and environmental interactions. Ocean nonlinear noise, whether induced by natural environmental changes [208,209] or human activities [210,211], may adversely affect ocean signal detection, presenting a complex challenge in this field. Ainslie and McColm [212] systematically explored the basic principles, techniques, and applications of underwater acoustics; and the impact of underwater noise on ecology and detection systems.

The application practice in ocean engineering has shown that ocean nonlinear noise affects ocean signal detection like sonar or radar. Underwater background noise is a complex issue that interferes with underwater target detection, causing signal distortion and attenuation. The potential for nonlinear noise to distort target echo signals and reduce the signal-to-noise ratio (SNR) makes it challenging to separate and identify the target signal. This challenge is particularly pronounced for sonar detection signals, as the interference of nonlinear noise affects the accuracy of target image feature extraction and recognition. The energy loss caused by nonlinear noise interference can lead to uneven signal image intensity, reducing the accuracy and efficiency of target detection.

Furthermore, the nonlinear noise generated by ocean intelligent detection devices and sensors themselves can affect their performance. Therefore, marine environmental noise prediction has become a valuable method for characterizing the detection performance of sonar systems [213]. The intrinsic causal relationship between wind speed and noise intensity can be used to quantify the impact of underwater background noise on marine sonar systems [214]. Da et al. [215] have pointed out that the nonlinear noise of the ocean determines the detection range of passive and active sonar. Similarly, Audoly & Lantéri [216] have highlighted that self-noise can reduce the detection performance of underwater sonar.

5.2. Coping Methods for Nonlinear Noise Impacts on Ocean Image Detection

5.2.1. Traditional Denoising Methods

Traditional underwater image denoising methods mainly rely on mathematical or physical models, such as spatial domain filtering, transform domain filtering, statistical modeling, and physical model-based methods.

Within the spatial domain filtering method, local or global operations are performed in the image pixel space to suppress noise through neighborhood pixel weighting or sorting. This method can be used for underwater sonar images or video denoising [217,218]. The industry standard for laser detection is spatial optical coherent filtering, a testament to this method's widespread use and reliability in reducing the interference of nonlinear noise in underwater laser detection and improving detection performance [219].

The transform domain method converts an image to the frequency domain or time-frequency domain and removes noise through threshold processing or coefficient correction. This includes wavelet threshold denoising [220,221] and Fourier Filtering [222]. The former is usually used for the removal of high-frequency noise from underwater blurred images, and the latter is usually used for elimination of band noise in satellite remote sensing images.

The statistical modeling methods are based on modeling the statistical characteristics of noise or images and restoring clean images through optimal estimation, which includes Wiener filtering [223] and Bayesian estimation [224]. The former is usually used for low-

frequency electronic noise suppression in sonar images, and the latter is usually used for color correction and noise joint optimization of underwater images.

The physics model-based approach refers to using ocean optical or acoustic propagation models to reverse engineer the process of noise formation. The key method is the underwater image restoration model, which estimates light rays' attenuation and backscattering components to restore clear images based on the Jaffe McGlamery equation. In this method, underwater observation images can be represented as

$$I(x) = J(x)e^{-\beta z} + B(1 - e^{-\beta z}) \quad (14)$$

where $I(x)$ and $J(x)$ represent the observation and clear images, respectively. B represents the background scattered light; β is the attenuation coefficient; and z is distance from the target object to the camera. Based on this, the underwater image restoration method can be represented as

$$J(x) = \frac{I(x) - B(1 - e^{-\beta z})}{e^{-\beta z}} \quad (15)$$

In practical applications, multiple traditional methods are often combined to process different types of noise in stages. Common processes are filtering denoising, physical model restoration, and wavelet denoising in sequence.

5.2.2. DL-Based Denoising Methods

However, traditional mathematical or physical denoising methods have limitations, such as insufficient modeling of complex noise, manual parameter tuning dependency, and image detail loss. Deep learning algorithms provide a more comprehensive and effective method to address the interference of nonlinear noise in ocean intelligent detection. They achieve automatic noise reduction and perform better in complex noise modeling and detail preservation.

Deep learning algorithms achieve efficient image denoising in ocean target detection and image recognition by automatically learning and processing the complex relationship between noise and effective signals. The specific noise reduction principle is mainly reflected in noise learning, signal reconstruction, and image enhancement. The first is noise residual or distribution learning with deep learning networks combined with data augmentation and transfer learning [225]. Second is the signal reconstruction with deep learning algorithms to enhance the signal features [226,227]. Last is distinguishing between noise and effective signals by automatically extracting multi-scale features using deep learning network structures. These principles are not singular and unchanging but tools we can shape according to our needs.

The first noise reduction principle is typically reflected in the commonly used deep learning methods such as denoising neural networks (DnCNNs) [228] and denoising diffusion probabilistic models (DDPMs) [229,230]. Moreover, the second noise reduction principle is typically reflected in the commonly used deep learning methods like denoising generative adversarial networks (DnGANs) [231,232]. DnCNN is an end-to-end denoising model based on a convolutional neural network (CNN), which directly learns the mapping from noisy images to clean images through multi-layer convolution and nonlinear activation functions. DDPM, intriguingly, belongs to the diffusion model, which simulates the forward process of gradually adding noise to clean images and then learns the backward denoising process. DnGAN achieves noise-to-clean image mapping through generator-discriminator adversarial training. Table 5 shows the core innovations, denoising strategies for dealing with complex marine environments, applicable scenarios, and limitations of the above three methods.

Table 5. Algorithm and application analysis of DnCNN, DDPM and DnGAN.

Method	Core Innovations	Denoising Strategies	Applicable Scenarios	Limitations
DnCNN	Residual Learning +Deep CNN	Directly learning noise distribution through residual mapping	Annotating static noise scenes with sufficient data	Weak dynamic noise processing and poor real-time performance
DDPM	Diffusion process +noise prediction	Modeling data distribution through Markov chain with progressive addition	Complex noise (pulse/biological noise) and unsupervised scenarios	High computational cost and insufficient physical consistency
DnGAN	Generator residual learning +multi-scale discriminator	Dynamic complex noise modeling and noise-clean image mapping	Multi-modal and dynamic noise (underwater optical, sonar, and remote noise)	High computational cost, insufficient data and physical consistency

The third noise reduction principle is usually realized by specific deep learning network structures with attention mechanisms or other image enhancement modules. This adaptable method reduces noise by enhancing and extracting multi-scale image features to achieve signal-to-noise separation. It is particularly useful in reducing the interference of nonlinear noise in underwater images derived from sonar or other sensors with intense noise and low resolutions. Table 6 shows related deep-learning network structures and denoising methods based on this principle of ocean detection applications.

Table 6. Deep learning network structures used for signal-to-noise separation and denoising.

References	Network Structures	Denoising Method
Li et al. [233]	Re-parameterization despeckling convolutional neural network (RepDNet)	Introduced pixel smoothing blocks (PSB) and edge enhancement blocks (EEB)
Zhou et al. [234]	Noise-aware deep learning model with fullband-subband attention network (NAFSA-Net)	Designing different subnetworks to estimate noise and signal components and extract signal features
Madhusudhanarao et al. [235]	Advanced recurrent neural network with novel loss function (ARRNN-NLF)	Enhanced Osprey Optimization Algorithm (EOOA) to enhance the denoising model
Domingos et al. [236]	VGGNet	Extracting multi-level features through deep convolutional stacking to identify noise
Lu et al. [237]	Deep blind despeckling network (DSPNet)	Introducing a feature pyramid network (FPN) and the atrous spatial pyramid pool (ASPP) to estimate and reduce random noise
Ji et al. [238]	Masked-Pre-training-Based Fast DIP (MPFDIP)	Improving denoising and performance by learning the intrinsic structural priors of images during the pre-training phase
Huo et al. [239]	High-Frequency Abrupt Signal Separation and Hybrid Attention Mechanism (HHDNet)	Utilizing a dual branch network architecture to handle high and low frequencies, and combines a hybrid attention module to remove high-frequency burst noise
Chen et al. [240]	Recurrent neural networks (RNNs)	Using a T-F-mask-aware bidirectional long short-term memory (Bi-LSTM) approach
Yao et al. [241]	Graph convolution networks (GCNs)	Extracting multi-domain features and deep features
Song and Shen [242]. Boosting R-CNN	Dual-Path Transformation Network (DPTN)	Constructing a neural network transformer based on feed-forward network to extract nonlinear noise features

6. Conclusions

This work reviews key technological developments in ocean intelligent perception and image processing. It covers ocean intelligent perception devices and image acquisition, image recognition and detection models, adaptive image processing processes, and coping methods for nonlinear noise interference. The review begins with a description of ocean intelligent detection technology and image acquisition via various sensors such as optical, acoustic, and laser sensors. It then shifts its focus to image processing network models, from the basic DCNN to one-stage and two-stage detection network models. The article provided a detailed analysis of the mathematical structures of the YOLO models and their improvements. It also reviewed adaptive image processing processes and their critical support for ocean image detection, such as image annotation, feature enhancement, and image segmentation. Theoretical research and practical application have shown that nonlinear noise from the seawater environment and the equipment significantly influence signal image processing. This work also focuses on countermeasures to cope with nonlinear noise impact, with deep learning methods for image feature enhancement, signal–noise separation, and noise suppression.

This review highlights that the most significant challenge stems from technological and environmental factors, presenting as underwater image quality degradation, annotated data scarcity, underwater sensor device resource limitations, and nonlinear noise interference. The solution method for underwater image quality degradation and ocean annotation data scarcity will focus on image enhancement, data generation, and transfer learning. Considering the device resource and environmental constraints, the optimization scheme based on segmentation technology mainly includes adaptive preprocessing enhancement and lightweight segmentation model design. On the one hand, it can introduce underwater image restoration networks as pre-segmentation modules to enhance the image data preprocessing capability of the detection model. On the other hand, it can use lightweight architectures to optimize the model by reducing the number of parameters through depth-wise separable convolution. Furthermore, it is necessary to combine nonlinear modeling, multi-modal data fusion, and physics-guided deep learning to develop more robust marine environmental noise suppression technologies in the future.

Author Contributions: Conceptualization, H.L., Y.L. and Y.T.; Methodology, H.L., Y.T. and T.Q.; Investigation, H.L. and Y.L.; Validation, H.L., Y.L. and T.Q.; Data curation, Y.L. and Y.T.; Writing—original draft preparation, Y.L. and T.Q.; Software, Y.L. and T.Q.; Formal analysis, T.Q. and Y.T.; Project administration, H.L. and Y.L.; Writing—review and editing, Y.L. and T.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Excellent Youth Program of Philosophy and Social Science of Anhui Universities (no. 2023AH030025); Graduate Education Innovation Fund of Anhui Polytechnic University (no. Xjky2022178, no. Xjky2022194); and Key Projects of Humanities and Social Sciences in Anhui Province’s Universities (no. 2022AH010459).

Data Availability Statement: The datasets supporting the conclusion of this article are included within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep learning
RUV	Remote underwater video
AUV	Autonomous underwater vehicle
DnCNN	denoising Convolutional Neural Network
DnGAN	denoising Generative Adversarial Network
YOLO	You Look Once
LiDAR	Light Detection and Ranging
Sonar	Sound Navigation and Ranging
FLS	Forward-looking sonar
SSS	Side-scan sonar
SAS	Synthetic aperture sonar
DIDSON	Dual Frequency Sonar
1DCTN	End-to-end underwater acoustic target recognition model
DCNN	Deep convolutional neural network
APRCNN	Audio Perspective Region-based Convolutional Neural Network
SVM	Support vector machine
R-CNN	Region-based Convolutional Neural Network
SSD	Single Shot MultiBox Detector
FCN	Fully Convolutional Network
MSE	Mean Squared Error
BCE	Binary Cross Entropy
PAN	Path aggregation networks
FPN	Feature pyramid networks
GPA	Gated path aggregate
SNR	Signal-to-noise ratio

References

1. Lu, F.Q.; Gao, X.Y.; Ma, J.; Xu, J.F.; Xue, Q.S.; Cao, D.S.; Quan, X.Q. Intelligent marine detection based on spectral imaging and neural network modeling. *Ocean. Eng.* **2024**, *310*, 118640. [\[CrossRef\]](#)
2. Yan, Y.J.; Liu, Y.D.; Fang, J.; Lu, Y.F.; Jiang, X.C. Application status and development trends for intelligent perception of distribution network. *High Volt.* **2021**, *6*, 938–954. [\[CrossRef\]](#)
3. Torres, A.; Abril, A.M.; Clua, E.E.G. A Time-Extended (24 h) Baited Remote Underwater Video (BRUV) for monitoring pelagic and nocturnal marine species. *J. Mar. Sci. Eng.* **2020**, *8*, 208. [\[CrossRef\]](#)
4. Wang, Y.Y.; Ma, X.R.; Wang, J.; Hou, S.L.; Dai, J.; Gu, D.B.; Wang, H.Y. Robust AUV visual loop-closure detection based on Variational Autoencoder Network. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8829–8838.
5. Lee, S.M.; Roh, M.I.; Jisang, H.; Lee, W. A method of estimating the locations of other ships from ocean images. *Korean J. Comput. Des. Eng.* **2020**, *25*, 320–328. [\[CrossRef\]](#)
6. Cheng, S.J.; Shi, X.C.; Mao, W.J.; Alkhalifah, T.A.; Yang, T.; Liu, Y.Z.; Sun, H.P. Elastic seismic imaging enhancement of sparse 4C ocean-bottom node data using deep learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5910214. [\[CrossRef\]](#)
7. Wang, C.; Stopa, J.E.; Vandemark, D.; Foster, R.; Ayet, A.; Mouche, A.; Chapron, B.; Sadowski, P. A multi-tagged SAR ocean image dataset identifying atmospheric boundary layer structure in winter tradewind conditions. *Geosci. Data J.* **2025**, *12*, e282. [\[CrossRef\]](#)
8. Dong, K.Y.; Liu, T.; Shi, Z.; Zhang, Y. Accurate and real-time visual detection algorithm for environmental perception of USVS under all-weather conditions. *J. Real-Time Image Process.* **2024**, *21*, 36. [\[CrossRef\]](#)
9. Zha, Z.J.; Ping, X.B.; Wang, S.L.; Wang, D.L. Deblurring of beamformed images in the ocean acoustic waveguide using deep learning-based deconvolution. *Remote Sens.* **2024**, *16*, 2411. [\[CrossRef\]](#)
10. Sattar, J.; Giguère, P.; Dudek, G. Sensor-based behavior control for an autonomous underwater vehicle. *Int. J. Robot. Res.* **2009**, *28*, 701–713.
11. Di Ciaccio, F. The supporting role of artificial intelligence and machine/deep learning in monitoring the marine environment: A bibliometric analysis. *Ecol. Quest.* **2024**, *35*, 1–30.

12. Xu, X.P.; Lin, X.X.; An, X.R. Introduction to image pro-processing subsystem of geostationary ocean color imager (GOCI). *Korean J. Remote Sens.* **2010**, *26*, 167–173.
13. Chen, S.D.; Liu, Y. Migration learning based on computer vision and its application in ocean image processing. *J. Coast. Res.* **2020**, *S104*, 281–285. [[CrossRef](#)]
14. Xue, L.Z.; Zeng, X.Y.; Jin, A.Q. A novel deep-learning method with channel attention mechanism for underwater target recognition. *Sensors* **2022**, *22*, 5492. [[CrossRef](#)] [[PubMed](#)]
15. Gong, W.J.; Tian, J.; Liu, J.Y. Underwater object classification method based on depthwise separable convolution feature fusion in sonar images. *Appl. Sci.* **2022**, *12*, 3268. [[CrossRef](#)]
16. Huang, Y.; Li, W.; Yuan, F. Speckle noise reduction in sonar imagebased on adaptive redundant dictionary. *J. Mar. Sci. Eng.* **2020**, *8*, 761.
17. Belcher, E.; Matsuyama, B.; Trimble, G. Object identification with acoustic lenses. In Proceedings of the Annual Conference of the Marine-Technology-Society, Honolulu, HI, USA, 5–8 November 2001.
18. Shin, Y.S.; Cho, Y.G.; Choi, H.T.; Kim, A. Comparative study of sonar image processing for underwater navigation. *J. Ocean. Eng. Technol.* **2016**, *30*, 214–220.
19. Chen, Y.H.; Liu, X.Y.; Jiang, J.Y.; Gao, S.Y.; Liu, Y.; Jiang, Y.Q. Reconstruction of degraded image transmitting through ocean turbulence via deep learning. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **2023**, *40*, 2215–2222. [[CrossRef](#)]
20. Yang, C.; Zhang, C.; Jiang, L.Y.; Zhang, X.W. Underwater image object detection based on multi-scale feature fusion. *Mach. Vis. Appl.* **2024**, *35*, 124. [[CrossRef](#)]
21. Wang, Z.; Guo, J.X.; Zeng, L.Y.; Zhang, C.L.; Wang, B.H. MLFFNet: Multilevel feature fusion network for object detection in sonar images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5119119.
22. Li, G.Y.; Liao, X.F.; Chao, B.H.; Jin, Y. Multi-scale feature fusion algorithm for underwater remote object detection in forward-looking sonar images. *J. Electron. Imaging* **2024**, *33*, 063031.
23. Xu, T.; Zhou, J.Y.; Guo, W.T.; Cai, L.; Ma, Y.K. Fine reconstruction of underwater images for environmental feature fusion. *Int. J. Adv. Robot. Syst.* **2021**, *18*, 17298814211039687. [[CrossRef](#)]
24. Gong, T.Y.; Zhang, M.M.; Zhou, Y.; Bai, H.H. Underwater image enhancement based on color feature fusion. *Electronics* **2023**, *12*, 4999. [[CrossRef](#)]
25. Chen, D.; Kang, F.; Li, J.J.; Zhu, S.S.; Liang, X.W. Enhancement of underwater dam crack images using multi-feature fusion. *Autom. Constr.* **2024**, *167*, 105727.
26. Zhu, Y.T.; Chao, X.P.; Wang, X.Q.; Chen, J.; Huang, H.F. An unsupervised ocean surface waves suppression algorithm based on sub-aperture SAR images. *Int. J. Remote Sens.* **2023**, *44*, 1460–1483.
27. Chao, X.P.; Wang, Q.S.; Wang, X.Q.; Chen, J.; Zhu, Y.T. Ocean-wave suppression for synthetic aperture radar images by depth counteraction method. *Remote Sens. Environ.* **2024**, *305*, 114086.
28. Jiang, Z.Y.; Gao, X.; Shi, L.; Li, N.; Zou, L. Detection of ocean internal waves based on modified deep convolutional generative adversarial network and WaveNet in moderate resolution imaging spectroradiometer images. *Appl. Sci.* **2023**, *13*, 11235. [[CrossRef](#)]
29. Liu, F.; Song, Q.Z.; Jin, G.H. Expansion of restricted sample for underwater acoustic signal based on Generative Adversarial Networks. In Proceedings of the 10th International Conference on Graphics and Image Processing, Chengdu, China, 12–14 December 2018.
30. Er, M.J.; Chen, M.J.; Zhang, Y.; Gao, W.X. Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. *Sensors* **2023**, *23*, 1990. [[CrossRef](#)]
31. Li, X.B.; Yan, L.; Qi, P.F.; Zhang, L.P.; Goudail, F.; Liu, T.G.; Zhai, J.S.; Hu, H.F. Polarimetric imaging via deep learning: A review. *Remote Sens.* **2023**, *15*, 1540. [[CrossRef](#)]
32. Li, J.; Xu, W.K.; Deng, L.M.; Xiao, Y.; Han, Z.Z.; Zheng, H.Y. Deep learning for visual recognition and detection of aquatic animals: A review. *Rev. Aquac.* **2023**, *15*, 409–433.
33. Saleh, A.; Sheaves, M.; Azghadi, M.R. Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish Fish.* **2022**, *23*, 977–999. [[CrossRef](#)]
34. Neupane, D.; Seok, J. A review on deep learning-based approaches for automatic sonar target recognition. *Electronics* **2020**, *9*, 1972. [[CrossRef](#)]
35. Whitaker, S.; Barnard, A.; Anderson, G.D.; Havens, T.C. Through-ice acoustic source tracking using vision transformers with ordinal classification. *Sensors* **2022**, *22*, 4703. [[CrossRef](#)]
36. Lei, Z.F.; Lei, X.F.; Na, W.; Zhang, Q.Y. Present status and challenges of underwater acoustic target recognition technology: A review. *Front. Phys.* **2022**, *10*, 1044890.
37. Kim, H.G.; Seo, J.-M.; Kim, S.M. Comparison of GAN deep learning methods for underwater optical image enhancement. *J. Ocean. Eng. Technol.* **2022**, *36*, 32–40. [[CrossRef](#)]

38. Chai, Y.Q.; Yu, H.H.; Xu, L.; Li, D.L.; Chen, Y.Y. Deep learning algorithms for sonar imagery analysis and its application in aquaculture: A review. *IEEE Sens. J.* **2023**, *23*, 28549–28563. [\[CrossRef\]](#)
39. Wang, C.Y.; Liao, H.Y.M. YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems. *Apsipa Trans. Signal Inf. Process.* **2024**, *13*, e29. [\[CrossRef\]](#)
40. Hussain, M. YOLOv1 to v8: Unveiling Each Variant-A Comprehensive Review of YOLO. *IEEE Access* **2024**, *12*, 42816–42833. [\[CrossRef\]](#)
41. Jiao, L.; Abdullah, M.I. YOLO series algorithms in object detection of unmanned aerial vehicles: A survey. *Serv. Oriented Comput. Appl.* **2024**, *18*, 269–298. [\[CrossRef\]](#)
42. Liu, S.; Xu, H.L.; Lin, Y.; Gao, L. Visual navigation for recovering an AUV by another AUV in shallow water. *Sensors* **2019**, *19*, 1889. [\[CrossRef\]](#)
43. Wang, X.M.; Zerr, B.; Thomas, H.; Clement, B.; Xie, Z.X. Pattern formation of multi-AUV systems with the optical sensor based on displacement-based formation control. *Int. J. Syst. Sci.* **2020**, *51*, 348–367.
44. Auster, P.J.; Lindholm, J.; Plourde, M.; Barber, K.; Singh, H. Camera configuration and use of AUVs to census mobile fauna. *Mar. Technol. Soc. J.* **2007**, *41*, 49–52.
45. Rao, J.J.; Xu, K.; Chen, J.B.; Lei, J.T.; Zhang, Z.; Zhang, Q.Y.; Giernacki, W.; Liu, M. Sea-surface target visual tracking with a multi-camera cooperation approach. *Sensors* **2022**, *22*, 693. [\[CrossRef\]](#)
46. Yang, D.F.; Solihin, M.I.; Zhao, Y.W.; Yao, B.C.; Chen, C.R.; Cai, B.Y.; Machmudah, A. A review of intelligent ship marine object detection based on RGB camera. *IET Image Process.* **2024**, *18*, 281–297.
47. Sahoo, A.; Dwivedy, S.K.; Robi, P. Advancements in the field of autonomous underwater vehicle. *Ocean. Eng.* **2019**, *181*, 145–160.
48. Alaie, H.K.; Farsi, H. Passive sonar target detection using statistical classifier and adaptive threshold. *Appl. Sci.* **2018**, *8*, 61. [\[CrossRef\]](#)
49. Abu, A.; Diamant, R. Enhanced fuzzy-based local informational algorithm for sonar image segmentation. *IEEE Trans. Image Process.* **2020**, *29*, 445–460.
50. Kim, S.H. A study on the position control system of the small ROV using sonar sensors. *J. Soc. Nav. Archit. Korea* **2008**, *45*, 579–589.
51. Jiang, J.J.; Wang, X.Q.; Duan, F.J.; Fu, X.; Huang, T.T.; Li, C.Y.; Ma, L.; Bu, L.R.; Sun, Z.B. A sonar-embedded disguised communication strategy by combining sonar waveforms and whale call pulses for underwater sensor platforms. *Appl. Acoust.* **2019**, *145*, 255–266.
52. Yuan, X.; Li, N.; Gong, X.B.; Yu, C.L.; Zhou, X.T.; Ortega, J.F.M. Underwater wireless sensor network-based delaunay triangulation (UWSN-DT) algorithm for sonar map fusion. *Comput. J.* **2023**, *67*, 1699–1709.
53. Xi, M.; Wang, Z.J.; He, J.Y.; Wang, Y.B.; Wen, J.B.; Xiao, S.; Yang, J.C. High-Precision underwater perception and path planning of AUVs based on Quantum-Enhanced. *IEEE Trans. Consum. Electron.* **2024**, *70*, 5607–5617.
54. Chen, G.J.; Cheng, D.G.; Chen, W.; Yang, X.; Guo, T.Z. Path planning for AUVs based on improved APF-AC algorithm. *CMC Comput. Mater. Contin.* **2024**, *78*, 3721–3741.
55. Jangir, P.K.; Ewans, K.C.; Young, I.R. On the functionality of radar and laser ocean wave sensors. *J. Mar. Sci. Eng.* **2022**, *10*, 1260. [\[CrossRef\]](#)
56. Kincade, K. Sensors and lasers map ebb and flow of ocean life. *Laser Focus World* **2003**, *39*, 91–96.
57. Laux, A.; Mullen, L.; Perez, P.; Zege, E. Underwater laser range finder. In *Ocean Sensing and Monitoring IV, Proceedings of the SPIE Defense, Security, and Sensing, Baltimore, MD, USA, 23–27 April 2012*; SPIE: Bellingham, WA, USA, 2012; Volume 8732, p. 83721B.
58. Fournier, G.R.; Bonnier, D.; Forand, J.L.; Pace, P.W. Range-gated underwater laser imaging-system. *Opt. Eng.* **1993**, *32*, 2185–2190.
59. Klepsvik, J.O.; Bjarnar, M.L. Laser-radar technology for underwater inspection, mapping. *Sea Technol.* **1996**, *37*, 49–53.
60. Long, H.; Shen, L.Q.; Wang, Z.Y.; Chen, J.B. Underwater forward-looking sonar images target detection via speckle reduction and scene prior. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5604413.
61. Franchi, M.; Ridolfi, A.; Allotta, B. Underwater navigation with 2D forward looking SONAR: An adaptive unscented Kalman filter-based strategy for AUVs. *J. Field Robot.* **2021**, *38*, 355–385.
62. Greene, A.; Rahman, A.F.; Kline, R.; Rahman, M.S. Side scan sonar: A cost-efficient alternative method for measuring seagrass cover in shallow environments. *Estuar. Coast. Shelf Sci.* **2018**, *207*, 250–258.
63. Belcher, E.; Hanot, W.; Burch, J. Dual-frequency identification sonar (DIDSON). In *Proceedings of the 2002 International Symposium on Underwater Technology, Tokyo, Japan, 16–19 April 2002*.
64. Joe, H.; Cho, H.; Sung, M.; Kim, J.; Yu, S.-c. Sensor fusion of two sonar devices for underwater 3D mapping with an AUV. *Auton. Robot.* **2021**, *45*, 543–560.
65. Christensen, J.H.; Mogensen, L.V.; Ravn, O. Side-Scan Sonar imaging: Real-time acoustic streaming. In *Proceedings of the 13th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles, Oldenburg, Germany, 22–24 September 2021*.
66. Zacchini, L.; Franchi, M.; Ridolfi, A. Sensor-driven autonomous underwater inspections: A receding-horizon RRT-based view planning solution for AUVs. *J. Field Robot.* **2022**, *39*, 499–527.

67. Li, X.H.; Li, Y.A.; Yu, J.; Chen, X.; Dai, M. PMHT approach for multi-target multi-sensor sonar tracking in clutter. *Sensors* **2015**, *15*, 28177–28192. [[CrossRef](#)]
68. Moursund, R.A.; Carlson, T.J.; Peters, R.D. A fisheries application of a dual-frequency identification sonar acoustic camera. In Proceedings of the ICES Symposium on Acoustics in Fisheries and Aquatic Ecology, Montpellier, France, 10–14 June 2002.
69. Able, K.W.; Grothues, T.M.; Rackovan, J.L.; Buderman, F.E. Application of Mobile Dual-frequency Identification Sonar (DIDSON) to Fish in Estuarine Habitats. *Northeast. Nat.* **2014**, *21*, 192–209.
70. Handegard, N.O.; Williams, K. Automated tracking of fish in trawls using the DIDSON (Dual frequency IDentification SONar). *ICES J. Mar. Sci.* **2008**, *65*, 636–644.
71. Nichols, O.C.; Eldredge, E.; Cadrin, S.X. Gray seal behavior in a fish weir observed using Dual-Frequency identification sonar. *Mar. Technol. Soc. J.* **2014**, *48*, 72–78.
72. McCann, E.L.; Johnson, N.S.; Hrodey, P.J.; Pangle, K.L. Characterization of sea lamprey stream entry using Dual-Frequency identification sonar. *Trans. Am. Fish. Soc.* **2018**, *147*, 514–524.
73. Zhao, W.; Li, X.; Pang, Z.Q.; Hao, C.P. A novel distributed bearing-only target tracking algorithm for underwater sensor networks with resource constraints. *IET Radar Sonar Navig.* **2024**, *18*, 1161–1177.
74. Tang, M.Q.; Ren, C.J.; Xin, Y.L. Efficient resource allocation algorithm for underwater wireless sensor networks based on improved stochastic gradient descent method. *AD HOC Sens. Wirel. Netw.* **2021**, *49*, 207–222.
75. Wu, H.J.; Wang, X.L.; Liao, H.B.; Jiao, X.B.; Liu, Y.Y.; Shu, X.J.; Wang, J.L.; Rao, Y.J. Signal processing in smart fiber-optic distributed acoustic sensor. *Acta Opt. Sin.* **2024**, *44*, 0106009.
76. Zhao, D.D.; Mao, W.B.; Chen, P.; Dang, Y.J.; Liang, R.H. FPGA-based real-time synchronous parallel system for underwater acoustic positioning and navigation. *IEEE Trans. Ind. Electron.* **2024**, *71*, 3199–3207.
77. Li, C.Y.; Guo, S.X. Characteristic evaluation via multi-sensor information fusion strategy for spherical underwater robots. *Inf. Fusion* **2023**, *95*, 199–214.
78. Periola, A.A.; Alonge, A.A.; Ogudo, K.A. Edge computing for big data processing in underwater applications. *Wirel. Netw.* **2022**, *28*, 2255–2271.
79. Fuentes, A.J.; Suchy, M.; Palomo, P.B. The greatest challenge for URN reduction in the oceans by means of engineering. In Proceedings of the MTS/IEEE Oceans Seattle Conference, Seattle, WA, USA, 27–31 October 2019.
80. Zheng, L.Y.; Liu, M.Q.; Zhang, S.L.; Liu, Z.A.; Dong, S.L. End-to-end multi-sensor fusion method based on deep reinforcement learning in UASNs. *Ocean. Eng.* **2024**, *305*, 117904.
81. Bhattacharjee, S.; Shanmugam, P.; Das, S. A deep-learning-based lightweight model for ship localizations in SAR Images. *IEEE Access* **2023**, *11*, 94415–94427.
82. Huang, J.F.; Zhang, T.J.; Zhao, S.J.; Zhang, L.; Zhou, Y.C. An underwater organism image dataset and a lightweight module designed for object detection networks. *ACM Trans. Multimed. Comput. Commun. Appl.* **2024**, *20*, 147.
83. Lu, Y.; Yang, M.; Liu, R.W. DSPNet: Deep learning-enabled blind reduction of speckle noise. In Proceedings of the 25th International Conference on Pattern Recognition, Electronic Network, Milan, Italy, 10–15 January 2021.
84. Wang, H.; Gao, N.; Xiao, Y.; Tang, Y. Image feature extraction based on improved FCN for UUV side-scan sonar. *Mar. Geophys. Res.* **2020**, *41*, 18.
85. Cheng, Z.; Huo, G.; Li, H. A multi-domain collaborative transfer learning method with multi-scale repeated attention mechanism for underwater side-scan sonar image classification. *Remote Sens.* **2022**, *14*, 355. [[CrossRef](#)]
86. Ribeiro, P.O.C.S.; dos Santos, M.M.; Drews, P.L.J.; Botelho, S.S.C.; Longaray, L.M.; Giacomo, G.G.; Pias, M.R. Underwater place recognition in unknown environments with triplet based acoustic image retrieval. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications, Orlando, FL, USA, 17–20 December 2018.
87. Wang, Z.; Guo, J.; Huang, W.; Zhang, S. Side-scan sonar image segmentation based on multi-channel fusion convolution neural networks. *IEEE Sens. J.* **2022**, *22*, 5911–5928.
88. Yu, S.B. Sonar image target detection based on deep learning. *Math. Probl. Eng.* **2022**, *2022*, 5294151.
89. Dai, Z.Z.; Liang, H.; Duan, T. Small-sample sonar image classification based on deep learning. *J. Mar. Sci. Eng.* **2022**, *10*, 1820. [[CrossRef](#)]
90. Ge, H.L.; Dai, Y.W.; Zhu, Z.Y.; Liu, R.B. A deep learning model applied to optical image target detection and recognition for the identification of underwater biostructures. *Machines* **2022**, *10*, 809. [[CrossRef](#)]
91. Zhang, Y.; Wang, X.W.; Sun, L.; Lei, P.S.; Chen, J.N.; He, J.; Zhou, Y.; Liu, Y.L. Mask-guided deep learning fishing net detection and recognition based on underwater range gated laser imaging. *Opt. Laser Technol.* **2024**, *171*, 110402. [[CrossRef](#)]
92. Yang, K.; Wang, B.; Fang, Z.D.; Cai, B.G. An end-to-end underwater acoustic target recognition model based on One-Dimensional Convolution and transformer. *J. Mar. Sci. Eng.* **2024**, *12*, 1793. [[CrossRef](#)]
93. Dong, K.Y.; Liu, T.; Shi, Z.; Zhang, Y. Visual detection algorithm for enhanced environmental perception of unmanned surface vehicles in complex marine environments. *J. Intell. Robot. Syst.* **2024**, *110*, 1. [[CrossRef](#)]

94. Li, L.L.; Zhang, S.J.; Wang, B. Plant disease detection and classification by deep learning—A review. *IEEE Access* **2021**, *9*, 56683–56698. [\[CrossRef\]](#)
95. Li, Y.; Tang, Y. Design on intelligent feature graphics based on convolution operation. *Mathematics* **2022**, *10*, 384. [\[CrossRef\]](#)
96. Singh, O.D.; Malik, A.; Yadav, V.; Gupta, S.; Dora, S. Deep segmenter system for recognition of micro cracks in solar cell. *Multimed. Tools Appl.* **2021**, *80*, 6509–6533. [\[CrossRef\]](#)
97. Liu, H.Y.; Li, Y. Interaction of Asymmetric Adaptive Network Structures and Parameter Balance in Image Feature Extraction and Recognition. *Symmetry* **2024**, *16*, 1651. [\[CrossRef\]](#)
98. Ma, Y.X.; Zhang, X.B.; Jiang, F.K.; Wei, Z.R.; Liu, C.G. Near-field geoacoustic inversion using bottom reflection signals via self-attention mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 10545–10558. [\[CrossRef\]](#)
99. Liu, Y.F.; Zhao, Y.J.; Gerstoft, P.; Zhou, F.; Qiao, G.; Yin, J.W. Deep transfer learning-based variable Doppler underwater acoustic communications. *J. Acoust. Soc. Am.* **2023**, *154*, 232–244. [\[CrossRef\]](#)
100. Scaradozzi, D.; de Marco, R.; Li Veli, D.; Lucchetti, A.; Screpanti, L.; Di Nardo, F. Convolutional Neural Networks for enhancing detection of Dolphin whistles in a dense acoustic environment. *IEEE Access* **2024**, *12*, 127141–127148.
101. Li, X.W.; Huang, W.M.; Peters, D.K.; Power, D. Assessment of synthetic aperture radar image preprocessing methods for iceberg and ship recognition with Convolutional Neural Networks. In Proceedings of the IEEE Radar Conference, Boston, MA, USA, 22–26 April 2019.
102. Ashok, P.; Latha, B. Feature extraction of underwater acoustic signal target using machine learning technique. *Trait. Du Signal* **2024**, *41*, 1303–1314.
103. Wang, H.Y.; Li, X.F. DeepBlue: Advanced convolutional neural network applications for ocean remote sensing. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 138–161. [\[CrossRef\]](#)
104. Uijlings, J.R.; van de Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171.
105. Yuan, Z.W.; Zhang, J. Feature extraction and image retrieval based on AlexNet. In Proceedings of the 8th International Conference on Digital Image Processing, Chengdu, China, 20–23 May 2016.
106. Liu, X.; Zhu, H.H.; Song, W.H.; Wang, J.H.; Yan, L.L.; Wang, K.L. Research on improved VGG-16 model based on transfer learning for acoustic image recognition of underwater search and rescue targets. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 18112–18128. [\[CrossRef\]](#)
107. Huang, X.; Li, N.; Pu, Y.Q.; Zhang, T.; Wang, B. Neuroprotective effects of ginseng phytochemicals: Recent perspectives. *Molecules* **2019**, *24*, 2939. [\[CrossRef\]](#)
108. Song, P.H.; Li, P.T.; Dai, L.H.; Wang, T.; Chen, Z. Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection. *Neurocomputing* **2023**, *530*, 150–164. [\[CrossRef\]](#)
109. Bao, S.D.; Meng, J.M.; Sun, L.N.; Liu, Y.X. Detection of ocean internal waves based on Faster R-CNN in SAR images. *J. Oceanol. Limnol.* **2020**, *38*, 55–63.
110. Zeng, L.C.; Sun, B.; Zhu, D.Q. Underwater target detection based on Faster R-CNN and adversarial occlusion network. *Eng. Appl. Artif. Intell.* **2021**, *100*, 104190.
111. Byeon, Y.; Kim, E.; Lim, H.J.; Kim, H.S. Development of a Faster R-CNN-based marine debris detection model for an embedded system. *J. Inst. Control. Robot. Syst.* **2021**, *27*, 1038–1043.
112. Faisal, M.; Chaudhury, S.; Sankaran, K.S.; Raghavendra, S.; Chitra, R.J.; Eswaran, M.; Boddu, R. Faster R-CNN algorithm for detection of plastic garbage in the ocean: A case for turtle preservation. *Math. Probl. Eng.* **2022**, *2022*, 3639222. [\[CrossRef\]](#)
113. Bi, W.H.; Jin, Y.; Li, J.X.; Sun, L.L.; Fu, G.W.; Jin, W. In-Situ detection method of Jellyfish based on improved Faster R-CNN and FP16. *IEEE Access* **2023**, *11*, 81803–81814.
114. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
115. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [\[CrossRef\]](#)
116. Zheng, Y.G.; Qi, K.T.; Zhang, H.S. Stripe segmentation of oceanic internal waves in synthetic aperture radar images based on Mask R-CNN. *Geocarto Int.* **2022**, *37*, 14480–14494. [\[CrossRef\]](#)
117. Qian, Y.; Liu, Q.; Zhu, H.M.; Fan, H.F.; Du, B.W.; Liu, S.C. Mask R-CNN for object detection in multitemporal SAR images. In Proceedings of the 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, Shanghai, China, 5–7 August 2019.
118. Jain, R.; Zaware, S.; Kacholia, N.; Bhalala, H.; Jagtap, O. Advancing underwater trash detection: Harnessing Mask R-CNN, YOLOv8, EfficientDet-D0 and YOLACT. In Proceedings of the 2nd International Conference on Sustainable Computing and Smart Systems, Coimbatore, India, 10–12 July 2024.
119. Conrady, C.R.; Er, S.; Attwood, C.G.; Roberson, L.A.; de Vos, L. Automated detection and classification of southern African Roman seabream using mask R-CNN. *Ecol. Inform.* **2022**, *69*, 101593. [\[CrossRef\]](#)

120. Lu, C.H.; Kong, Y.; Guan, Z.Y. A mask R-CNN model for reidentifying extratropical cyclones based on quasi-supervised thought. *Sci. Rep.* **2020**, *10*, 15011. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Liu, C.; Wang, Z.; Wang, S.; Tang, T.; Tao, Y.; Yang, C.; Li, H.; Liu, X.; Fan, X. A new dataset, poisson GAN and AquaNet for underwater object grabbing. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 2831–2844. [\[CrossRef\]](#)
122. Mana, S.C.; Sasipraba, T. An intelligent deep learning enabled marine fish species detection and classification model. *Int. J. Artif. Intell. Tools* **2022**, *31*, 2250017. [\[CrossRef\]](#)
123. Bi, X.L.; Xiao, B.; Li, W.S.; Gao, X.B. IEMask R-CNN: Information-Enhanced Mask R-CNN. *IEEE Trans. Big Data* **2023**, *9*, 688–700. [\[CrossRef\]](#)
124. Lee, H.; Eum, S.; Kwon, H. ME R-CNN: Multi-Expert R-CNN for object detection. *IEEE Trans. Image Process.* **2020**, *29*, 1030–1044. [\[CrossRef\]](#)
125. Liu, Y.K.; Huang, Y.W.; Yin, Y.L. SE-Mask R-CNN: An improved Mask R-CNN for apple detection and segmentation. *J. Intell. Fuzzy Syst.* **2021**, *41*, 6715–6725. [\[CrossRef\]](#)
126. Xu, F.J.; Huang, J.X.; Wu, J.; Jiang, L.Y. Active Mask-Box Scoring R-CNN for sonar image instance segmentation. *Electronics* **2022**, *11*, 2048. [\[CrossRef\]](#)
127. Yi, D.W.; Ahmedov, H.B.; Jiang, S.Y.; Li, Y.R.; Flinn, S.J.; Fernandes, P.G. Coordinate-Aware Mask R-CNN with Group Normalization: A underwater marine animal instance segmentation framework. *Neurocomputing* **2024**, *583*, 127488.
128. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016.
129. Yan, C.Q.; Zhang, H.; Li, X.L.; Yuan, D. R-SSD: Refined single shot multibox detector for pedestrian detection. *Appl. Intell.* **2022**, *52*, 10430–10447.
130. Wang, Z.H.; Yang, S.; Shi, M.J.; Qin, K.Y. FDA-SSD: Fast Depth-Assisted Single-Shot MultiBox Detector for 3D tracking based on monocular vision. *Appl. Sci.* **2022**, *12*, 1164. [\[CrossRef\]](#)
131. Wang, L.Y.; Wang, X.W.; Li, B. Data-driven model SSD-BSP for multi-target coal-gangue detection. *Measurement* **2023**, *219*, 113244.
132. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–2 July 2016.
133. Chen, X.Q.; Wang, M.L.; Ling, J.; Wu, H.F.; Wu, B.; Li, C.F. Ship imaging trajectory extraction via an aggregated you only look once (YOLO) model. *Eng. Appl. Artif. Intell.* **2024**, *130*, 107742. [\[CrossRef\]](#)
134. Shi, Y.G.; Li, S.K.; Liu, Z.Y.; Zhou, Z.G.; Zhou, X.H. MTP-YOLO: You only look once based maritime tiny person detector for emergency rescue. *J. Mar. Sci. Eng.* **2024**, *12*, 669. [\[CrossRef\]](#)
135. Altarez, R.D. Faster R-CNN, RetinaNet and Single Shot Detector in different ResNet backbones for marine vessel detection using cross polarization C-band SAR imagery. *Remote Sens. Appl. -Soc. Environ.* **2024**, *36*, 101297.
136. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
137. Zhou, L.M.; Rao, X.H.; Li, Y.H.; Zuo, X.Y.; Qiao, B.J.; Lin, Y.H. A Lightweight object detection method in aerial images based on dense feature fusion path aggregation network. *ISPRS Int. J. Geo Inf.* **2022**, *11*, 189.
138. Han, M. E-Bayesian estimation and its E-MSE under the scaled squared error loss function, for exponential distribution as example. *Commun. Statics-Simul. Comput.* **2019**, *48*, 1880–1890.
139. Ting, A.; Santos, J.; Gultinan, E.; Gultinan, E. Using machine learning to predict multiphase flow through complex fractures. *Energies* **2022**, *15*, 8871. [\[CrossRef\]](#)
140. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning what you want to learn using programmable gradient information. In Proceedings of the 18th European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024.
141. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-time end-to-end object detection. *arXiv* **2024**, arXiv:2405.14458.
142. Redmon, J.; Farhadi, A. Yolo9000: Better, faster, stronger. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
143. Shafiee, M.J.; Chywl, B.; Li, F.; Wong, A. Fast YOLO: A fast You Only Look Once system for real-time embedded object detection in video. *arXiv* **2017**, arXiv:1709.05943.
144. Panchal, V.; Sankla, H.; Sharma, A.C.S.S. FPGA implementation of proposed number plate localization algorithm based on YOLOv2 (You Only Look Once). *Microsyst. Technol.* **2023**, *10*, 1501–1513.
145. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
146. Gunawan, C.R.; Nurdin, N.; Fajriana, F. Design of a real-time object detection prototype system with YOLOv3 (You Only Look Once). *Int. J. Eng. Sci. Inf. Technol.* **2022**, *2*, 96–99.
147. Choi, J.; Chun, D.; Kim, H.; Lee, H.J. Gaussian YOLOv3: An accurate and fast object detector using localization uncertainty for autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.

148. Bochkovskiy, A.; Wang, C.-Y.; Liao, H. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
149. Liu, P.; Wang, Q. SCA-YOLOv4: You only look once with squeeze-and-excitation, coordinate attention and adaptively spatial feature fusion. *Signal Image Video Process.* **2024**, *18*, 7093–7106.
150. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-YOLOv4: Scaling cross stage partial network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Los Angeles, CA, USA, 19–25 June 2021.
151. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
152. Zendehele, N.; Chen, H.; Leu, M.C. Real-time tool detection in smart manufacturing using You-Only-Look Once (YOLO)v5. *Manuf. Lett.* **2023**, *35*, 1052–1059.
153. Gupta, C.; Gill, N.S.; Gulia, P.; Chatterjee, J.M. A novel finetuned YOLOv6 transfer learning model for real-time object detection. *J. Real-Time Image Process.* **2023**, *20*, 42.
154. Li, N.; Wang, M.L.; Yang, G.C.; Li, B.; Yuan, B.H.; Xu, S.K. DENS-YOLOv6: A small object detection model for garbage detection on water surface. *Multimed. Tools Appl.* **2023**, *83*, 55751–55771. [[CrossRef](#)]
155. Wang, J.; Li, Q.Q.; Fang, Z.Q.; Zhou, X.L.; Tang, Z.W.; Han, Y.L.; Ma, Z.L. YOLOv6-ESG: A lightweight seafood detection method. *J. Mar. Sci. Eng.* **2023**, *11*, 1623. [[CrossRef](#)]
156. Cai, L.M.; Zha, G.Z.; Lin, M.S.; Wang, X.; Zhang, H.H. Ocean internal wave detection in SAR images based on improved YOLOv7. *IEEE Access* **2024**, *12*, 146852–146865.
157. Jiang, Z.K.; Su, L.; Sun, Y.X. YOLOv7-Ship: A lightweight algorithm for ship object detection in complex marine environments. *J. Mar. Sci. Eng.* **2024**, *12*, 190. [[CrossRef](#)]
158. Patel, K.; Bhatt, C.; Mazzeo, P.L. Improved ship detection algorithm from satellite images using YOLOv7 and Graph Neural Network. *Algorithms* **2022**, *15*, 473. [[CrossRef](#)]
159. Song, G.W.; Chen, W.; Zhou, Q.L.; Guo, C.K. Underwater robot target detection algorithm based on YOLOv8. *Electronics* **2024**, *13*, 3374. [[CrossRef](#)]
160. Zhang, F.B.; Cao, W.Y.; Gao, J.; Liu, S.B.; Li, C.Y.; Song, K.; Wang, H.W. Underwater object detection algorithm based on an improved YOLOv8. *J. Mar. Sci. Eng.* **2024**, *12*, 1991. [[CrossRef](#)]
161. Qu, S.M.; Cui, C.; Duan, J.L.; Lu, Y.Y.; Pang, Z.L. Underwater small target detection under YOLOv8-LA model. *Sci. Rep.* **2024**, *14*, 16108. [[CrossRef](#)]
162. Sharma, A.; Kumar, V.; Longchamps, L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agric. Technol.* **2024**, *9*, 100648.
163. Yuan, X.; Li, J.P.; Wang, W.W.; Zhou, X.T.; Li, N.; Yu, C.L. Improved YOLOv9 for underwater side scan sonar target detection. *Comput. J.* **2024**, *2024*, bxae134. [[CrossRef](#)]
164. Tu, W.; Yu, H.; Wu, Z.J.; Li, J.; Cui, Z.B.; Yang, Z.Y.; Zhang, X.; Wang, Y. YOLOv10-UDFishNet: Detection of diseased Takifugu rubripes juveniles in turbid underwater environments. *Aquac. Int.* **2025**, *33*, 125.
165. Gomes-Pereira, J.N.; Auger, V.; Beisiegel, K.; Benjamin, R.; Bergmann, M.; Bowden, D.; Buhl-Mortensen, P.; De Leo, F.C.; Dionísio, G.; Durden, J.M.; et al. Current and future trends in marine image annotation software. *Prog. Oceanogr.* **2016**, *149*, 106–120. [[CrossRef](#)]
166. Schoening, T.; Osterloff, J.; Nattkemper, T.W. RecoMIA-recommendations for marine image annotation: Lessons learned and future directions. *Front. Mar. Sci.* **2016**, *3*, 59. [[CrossRef](#)]
167. Giordano, D.; Palazzo, S.; Spampinato, C. A diversity-based search approach to support annotation of a large fish image dataset. *Multimed. Syst.* **2016**, *22*, 725–736.
168. Sánchez, J.S.; Lisani, J.L.; Catalán, I.A.; Alvarez-Ellacuría, A. Leveraging bounding box annotations for fish segmentation in underwater images. *IEEE Access* **2023**, *11*, 125984–125994.
169. Jeon, M.H.; Lee, Y.; Shin, Y.S.; Jang, H.; Yeu, T.K.; Kim, A. Synthesizing image and automated annotation tool for CNN based under water object detection. *J. Korea Robot. Soc.* **2019**, *14*, 139–149.
170. Zurowietz, M.; Langenkämper, D.; Hosking, B.; Ruhl, H.A.; Nattkemper, T.W. MAIA-A machine learning assisted image annotation method for environmental monitoring and exploration. *PLoS ONE* **2018**, *13*, e0207498.
171. Wu, Q.H.; Liu, Y.; Zhang, J.Y.; Wang, Y.P. Intelligent annotation algorithm based on deep-sea macrobenthic images. *Comput. Inform.* **2022**, *41*, 739–756.
172. Lisani, J.L.; Petro, A.B.; Sbert, C.; Alvarez-Ellacuría, A.; Catalan, I.A.; Palmer, M. Analysis of underwater image processing methods for annotation in deep learning based fish detection. *IEEE Access* **2022**, *10*, 130359–130372.
173. Li, Z.C.; Xie, H.J.; Feng, J.Y.; Wang, Z.B.; Yuan, Z.Z. YOLOv7-PE: A Precise and Efficient Enhancement of YOLOv7 for Underwater Target Detection. *IEEE Access* **2024**, *12*, 133937–133951.
174. Ren, L.Q.; Li, Z.Y.; He, X.Y.; Kong, L.Y.; Zhang, Y.H. An underwater target detection algorithm based on attention mechanism and improved YOLOv7. *CMC Comput. Mater. Contin.* **2024**, *78*, 2829–2845.
175. Chen, X.; Yuan, M.J.H.; Yang, Q.; Yao, H.Y.; Wang, H.Y. Underwater-YCC: Underwater target detection optimization algorithm based on YOLOv7. *J. Mar. Sci. Eng.* **2023**, *11*, 995. [[CrossRef](#)]

176. Yang, Q.; Meng, H.J.; Gao, Y.C.; Gao, D.X. A real-time object detection method for underwater complex environments based on FasterNet-YOLOv7. *J. Real Time Image Process.* **2023**, *21*, 8.
177. Liu, K.Y.; Sun, Q.; Sun, D.M.; Peng, L.; Yang, M.D.; Wang, N.Z. Underwater target detection based on improved YOLOv7. *J. Mar. Sci. Eng.* **2023**, *11*, 677. [\[CrossRef\]](#)
178. Qin, J.H.; Zhou, H.L.; Yi, H.; Ma, L.Y.; Nie, J.H.; Huang, T.T. YOLOv7-GCM: A detection algorithm for creek waste based on improved YOLOv7 model. *Pattern Anal. Appl.* **2024**, *27*, 116.
179. Yu, G.Y.; Cai, R.L.; Su, J.P.; Hou, M.X.; Deng, R.L. U-YOLOv7: A network for underwater organism detection. *Ecol. Inform.* **2023**, *75*, 102108.
180. Fu, J.S.; Tian, Y. Improved YOLOv7 underwater object detection based on attention mechanism. *Eng. Lett.* **2024**, *32*, 1377–1384.
181. Gao, Y.; Li, Z.Y.; Zhang, K.Y.; Kong, L.Y. GCP-YOLO: A lightweight underwater object detection model based on YOLOv7. *J. Real-Time Image Process.* **2024**, *22*, 3.
182. Zhao, L.; Yun, Q.; Yuan, F.C.; Ren, X.; Jin, J.W.; Zhu, X.C. YOLOv7-CHS: An emerging model for underwater object detection. *J. Mar. Sci. Eng.* **2023**, *11*, 1949. [\[CrossRef\]](#)
183. Yi, W.G.; Wang, B. Research on underwater small target detection algorithm based on improved YOLOv7. *IEEE Access* **2023**, *11*, 66818–66827. [\[CrossRef\]](#)
184. Lu, D.H.; Yi, J.X.; Wang, J. Enhanced YOLOv7 for improved underwater target detection. *J. Mar. Sci. Eng.* **2024**, *12*, 1127. [\[CrossRef\]](#)
185. Zhao, M.; Zhou, H.B.; Li, X. YOLOv7-SN: Underwater target detection algorithm based on improved YOLOv7. *Symmetry* **2024**, *16*, 514. [\[CrossRef\]](#)
186. Chen, Z.; Xie, G.H.; Chen, M.S.; Qiu, H.B. Model for underwater acoustic target recognition with attention mechanism based on residual concatenate. *J. Mar. Sci. Eng.* **2024**, *12*, 24. [\[CrossRef\]](#)
187. Ou, J.Y.; Shen, Y.J. Underwater target detection based on improved YOLOv7 algorithm with BiFusion neck structure and MPDIoU loss function. *IEEE Access* **2024**, *12*, 105165–105177. [\[CrossRef\]](#)
188. Yuan, M.; Meng, H.; Wu, J.B.; Cai, S.W. Global Recurrent Mask R-CNN: Marine ship instance segmentation. *Comput. Graph.* **2025**, *126*, 104112. [\[CrossRef\]](#)
189. Zeng, J.X.; Ouyang, H.; Liu, M.; Leng, L.; Fu, X. Multi-scale YOLACT for instance segmentation. *J. King Saud Univ.—Comput. Inf. Sci.* **2022**, *34*, 9419–9427. [\[CrossRef\]](#)
190. Huang, H.H.; Zuo, Z.; Sun, B.; Wu, P.; Zhang, J.J. DSA-SOLO: Double Split Attention SOLO for side-scan sonar target segmentation. *Appl. Sci.* **2022**, *12*, 9365. [\[CrossRef\]](#)
191. Song, Y.; Zhu, Y.; Li, G.; Feng, C.; He, B.; Yan, T. Side scan sonar segmentation using deep convolutional neural network. In Proceedings of the Oceans, Anchorage, AK, USA, 18–21 September 2017.
192. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
193. Chaurasia, K.; Nandy, R.; Pawar, O.; Singh, R.R.; Ahire, M. Semantic segmentation of high-resolution satellite images using deep learning. *Earth Sci. Inform.* **2021**, *14*, 2161–2170. [\[CrossRef\]](#)
194. Zheng, Y.G.; Zhang, H.S.; Qi, K.T.; Ding, L.Y. Stripe segmentation of oceanic internal waves in SAR images based on SegNet. *Geocarto Int.* **2022**, *37*, 8567–8578. [\[CrossRef\]](#)
195. Yu, H.F.; Li, X.B.; Feng, Y.K.; Han, S. Multiple attentional path aggregation network for marine object detection. *Appl. Intell.* **2022**, *53*, 2434–2451. [\[CrossRef\]](#)
196. Li, B.; Xu, J.; Chu, L.L.; Yang, Y.Q.; Huang, X.L.; Liu, P. Oil film semantic segmentation method in X-band marine radar remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1503205. [\[CrossRef\]](#)
197. Han, K.B.; Chen, M.J.; Gao, C.Z.; Qing, C.M. DCA-Unet: Enhancing small object segmentation in hyperspectral images with Dual Channel Attention Unet. *J. Frankl. Inst.* **2025**, *362*, 107532. [\[CrossRef\]](#)
198. Li, J.J.; Wang, H.F.; Zhang, A.B.; Liu, Y.L. Semantic segmentation of hyperspectral remote sensing images based on PSE-UNet model. *Sensors* **2022**, *22*, 9678. [\[CrossRef\]](#)
199. Hasimoto-Beltran, R.; Canul-Ku, M.; Mendez, G.M.D.; Ocampo-Torres, F.J.; Esquivel-Trava, B. Ocean oil spill detection from SAR images based on multi-channel deep learning semantic segmentation. *Mar. Pollut. Bull.* **2023**, *188*, 114651. [\[CrossRef\]](#)
200. Miao, J.; Xu, S.W.; Zou, B.X.; Qiao, Y.H. ResNet based on feature-inspired gating strategy. *Multimed. Tools Appl.* **2022**, *81*, 19283–19300. [\[CrossRef\]](#)
201. Chen, H.T.; Lee, S.; Yao, D.; Jeong, D. Sea clutter image segmentation method of high frequency surface wave radar based on the improved Deeplab network. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2022**, *E105A*, 730–733. [\[CrossRef\]](#)
202. Liu, F.F.; Fang, M. Semantic segmentation of underwater images based on improved Deeplab. *J. Mar. Sci. Eng.* **2020**, *8*, 188. [\[CrossRef\]](#)
203. Cheng, L.X.; Li, Y.; Zhao, K.J.; Liu, B.X.; Sun, Y.H. A two-stage oil spill detection method based on an improved superpixel module and DeepLab V3+ using SAR images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 3508020. [\[CrossRef\]](#)

204. Yu, Z.M.; Wan, F.; Lei, G.B.; Xiong, Y.; Xu, L.; Ye, Z.W.; Liu, W.; Zhou, W.; Xu, C.Z. RSLC-Deeplab: A ground object classification method for high-resolution remote sensing images. *Electronics* **2023**, *12*, 3653. [\[CrossRef\]](#)
205. Chen, Y.T.; Li, Y.Y.; Wang, J.S. An end-to-end oil-spill monitoring method for multisensory satellite images based on deep semantic segmentation. *Sensors* **2020**, *20*, 725. [\[CrossRef\]](#) [\[PubMed\]](#)
206. Wang, Q.; Zhang, Y.X.; He, B. Intelligent marine survey: Lightweight multi-scale attention adaptive segmentation framework for underwater target detection of AUV. *IEEE Trans. Autom. Sci. Eng.* **2024**, *22*, 1913–1927. [\[CrossRef\]](#)
207. Brockett, P.L.; Hinich, M.; Wilson, G.R. Nonlinear and non-Gaussian ocean noise. *J. Acoust. Soc. Am.* **1987**, *82*, 1386–1394. [\[CrossRef\]](#)
208. Williams, R.; Veirs, S.; Veirs, V.; Ashe, E.; Mastick, N. Approaches to reduce noise from ships operating in important killer whale habitats. *Mar. Pollut. Bull.* **2019**, *139*, 459–469.
209. Qian, T.; Li, Y. Nonlinear perception characteristics analysis of ocean white noise based on deep learning algorithms. *Mathematics* **2024**, *12*, 2892. [\[CrossRef\]](#)
210. Matthews, L.P.; Parks, S.E. An overview of North Atlantic right whale acoustic behavior, hearing capabilities, and responses to sound. *Mar. Pollut. Bull.* **2021**, *173*, 113043. [\[PubMed\]](#)
211. Traverso, F.; Vernazza, G.; Trucco, A. Simulation of non-White and non-Gaussian underwater ambient noise. In Proceedings of the Oceans MTS/IEEE, Yeosu, Republic of Korea, 21–24 May 2012.
212. Ainslie, M.A.; McColm, A.C.E. *Underwater Acoustics: Principles and Techniques*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–10.
213. Rogers, J.S.; Wales, S.C.; Means, S.L. Ambient noise forecasting with a large acoustic array in a complex shallow water environment. *J. Acoust. Soc. Am.* **2017**, *142*, EL473–EL477.
214. Wilson, J.D.; Makris, N.C. Quantifying hurricane destructive power, wind speed, and air-sea material exchange with natural undersea sound. *Geophys. Res. Lett.* **2008**, *35*, L10603.
215. Da, L.L.; Wang, C.; Han, M.; Zhang, L. Ambient noise spectral properties in the north area of Xisha. *Acta Oceanol. Sin.* **2014**, *33*, 206–211.
216. Audoly, C.; Lantéri, F. Modeling of the influence of the self noise on the array gain of a sonar system. *Appl. Acoust.* **2001**, *62*, 1095–1105.
217. Zang, X.; Yin, T.; Hou, Z.; Mueller, R.P.; Deng, Z.D.; Jacobson, P.T. Deep learning for automated detection and identification of migrating American eel *Anguilla rostrata* from imaging sonar data. *Remote Sens.* **2021**, *13*, 2671. [\[CrossRef\]](#)
218. Xiao, J.; Zou, W.; Zhang, S.; Lei, J.; Wang, W.; Wang, Y. Video denoising algorithm based on improved dual-domain filtering and 3D block matching. *IET Image Process.* **2018**, *12*, 2250–2257.
219. Jantzi, A.; Rumbaugh, L.; Jemison, W. Spatial optical coherence filtering for scatter rejection in underwater laser systems. In Proceedings of the Ocean Sensing and Monitoring XI, Baltimore, MD, USA, 16–17 April 2019.
220. Zhang, Y.Y.; Yang, Z.X.; Du, X.L.; Luo, X.Y. A new method for denoising underwater acoustic signals based on EEMD, correlation coefficient, permutation entropy, and wavelet threshold denoising. *J. Mar. Sci. Appl.* **2024**, *23*, 101–112.
221. Fan, C.Y.; Li, C.F.; Yang, S.H.; Liu, X.Y.; Liao, Y.Q. Application of CEEMDAN combined wavelet threshold denoising algorithm to suppressing scattering cluster in underwater lidar. *Acta Physic Sin.* **2023**, *72*, 224203.
222. Yan, L.; Piao, S.C.; Xu, F. Orthogonal waveform separation in multiple-input and multiple-output imaging sonar with fractional Fourier filtering. *Iet Radar Sonar Navig.* **2021**, *15*, 471–484.
223. Hong, J.; Bae, I.; Seok, J. Wiener filtering-based ambient noise reduction technique for improved acoustic target detection of directional frequency analysis and recording sonobuoy. *J. Acoust. Soc. Korea* **2022**, *41*, 192–198.
224. Hubert, P.; Padovese, L.; Martins, F. Bayesian estimation of the size distribution of air bubbles in water. *J. Braz. Soc. Mech. Sci. Eng.* **2021**, *43*, 527.
225. Yang, W.Y.; Chang, W.L.; Song, Z.C.; Niu, F.Q.; Wang, X.Y.; Zhang, Y. denoising odontocete echolocation clicks using a hybrid model with convolutional neural network and long short-term memory network. *J. Acoust. Soc. Am.* **2023**, *154*, 938–947.
226. Wang, S.W.; Song, P.; Tan, J.; He, B.S.; Xia, D.M.; Wang, Q.Q.; Du, G.N. Deep learning-based attenuation for shear-wave leakage from ocean-bottom node data. *Geophysics* **2023**, *88*, V127–V137.
227. Sorkhabi, O.M.; Asgari, J.; Amiri-Simkooei, A. Monitoring of Caspian Sea-level changes using deep learning-based 3D reconstruction of GRACE signal. *Measurement* **2021**, *174*, 109004. [\[CrossRef\]](#)
228. Amiri, A.; Kimiaghalam, B. Robust watermarking with PSO and DnCNN. *Signal Image Video Process.* **2024**, *18* (Suppl. S1), 663–676. [\[CrossRef\]](#)
229. Lu, S.Q.; Guan, F.X.; Zhang, H.Y.; Lai, H.T. Underwater image enhancement method based on denoising diffusion probabilistic model. *J. Vis. Commun. Image Represent.* **2023**, *96*, 103926. [\[CrossRef\]](#)
230. Guan, M.S.; Xu, H.Y.; Jiang, G.Y.; Yu, M.; Chen, Y.Y.; Luo, T.; Zhang, X.B. DiffWater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 2319–2335. [\[CrossRef\]](#)

231. Zheng, Z.Q.; Wang, C.; Yu, Z.B.; Zheng, H.Y.; Zheng, B. Instance map based image synthesis with a denoising generative adversarial network. *IEEE Access* **2018**, *6*, 33654–33665. [[CrossRef](#)]
232. Saranya, S.; Vellaturi, P.K.; Velichala, V.R.; Vemule, C.K. Analyzing image denoising using generative adversarial network. *J. Pharm. Negat. Results* **2022**, *13* (Suppl. S3), 307–310.
233. Li, Z.Y.; Wang, Z.S.; Chen, D.S.; Yip, T.L.; Teixeira, A.P. RepDNet: A re-parameterization despeckling network for autonomous underwater side-scan sonar imaging with prior-knowledge customized convolution. *Def. Technol.* **2024**, *35*, 259–274. [[CrossRef](#)]
234. Zhou, A.L.; Zhang, W.; Li, X.Y.; Xu, G.J.; Zhang, B.B.; Ma, Y.X.; Song, J.Q. A novel noise-aware deep learning model for underwater acoustic denoising. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4202813. [[CrossRef](#)]
235. Madhusudhanarao, K.; Krishna, K.M.; Krishna, B.T. An efficient noise reduction technique in underwater acoustic signals using enhanced optimization-based residual recurrent neural network with novel loss function. *Int. J. Wavelets Multiresolution Inf. Process.* **2024**, *23*, 2450048. [[CrossRef](#)]
236. Domingos, L.C.F.; Santos, P.E.; Skelton, P.S.M.; Brinkworth, R.S.A.; Sammut, K. An investigation of preprocessing filters and deep learning methods for vessel type classification with underwater acoustic data. *IEEE Access* **2022**, *10*, 117582–117596. [[CrossRef](#)]
237. Lu, Y.; Liu, R.W.; Chen, F.; Xie, L. Learning a deep convolutional network for speckle noise reduction in underwater sonar images. In Proceedings of the 11th International Conference on Machine Learning and Computing, Zhuhai, China, 22–24 February 2019.
238. Ji, S.; Xu, S.; Xu, S.; Cheng, Q.; Xiao, N.; Zhou, C.; Xiong, M. A masked-pre-training-based fast deep image prior denoising model. *Appl. Sci.* **2024**, *14*, 5125. [[CrossRef](#)]
239. Huo, C.L.; Zhang, D.; Yang, H.Y. An underwater image denoising method based on high-frequency abrupt signal separation and hybrid attention mechanism. *Sensors* **2024**, *24*, 4578. [[CrossRef](#)] [[PubMed](#)]
240. Chen, J.; Liu, C.; Xie, J.W.; An, J.; Huang, N. Time-Frequency Mask-Aware Bidirectional LSTM: A deep learning approach for underwater acoustic signal separation. *Sensors* **2022**, *22*, 5598. [[CrossRef](#)] [[PubMed](#)]
241. Yao, X.H.; Yang, H.H.; Sheng, M.P. Feature fusion based on Graph Convolution Network for modulation classification in underwater communication. *Entropy* **2023**, *25*, 1096. [[CrossRef](#)]
242. Song, Y.Q.; Liu, F.; Shen, T.S. Method of underwater acoustic signal denoising based on dual-path transformer network. *IEEE Access* **2024**, *12*, 81483–81494.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.