remote sensing

MDPI

Article

AquaPile-YOLO: Pioneering Underwater Pile Foundation Detection with Forward-Looking Sonar Image Processing

Zhongwei Xu [1,2], Rui Wang [1,*], Tianyu Cao [3], Wenbo Guo [3], Bo Shi [3] and Qiqi Ge [3]

[1] Department of Information and Communication Engineering, Tongji University, Shanghai 201804, China; 2180135@tongji.edu.cn
[2] China State Shipbuilding Corporation Haiying Enterprise Group Co., Ltd., Wuxi 214061, China
[3] Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China; caotianyu2023@sjtu.edu.cn (T.C.); wenbo.guo@sjtu.edu.cn (W.G.); tommy7080@sjtu.edu.cn (B.S.); gqq@sjtu.edu.cn (Q.G.)
* Correspondence: ruiwang@tongji.edu.cn

Abstract: Underwater pile foundation detection is crucial for environmental monitoring and marine engineering. Traditional methods for detecting underwater pile foundations are labor-intensive and inefficient. Deep learning-based image processing has revolutionized detection, enabling identification through sonar imagery analysis. This study proposes an innovative methodology, named the AquaPile-YOLO algorithm, for underwater pile foundation detection. Our approach significantly enhances detection accuracy and robustness by integrating multi-scale feature fusion, improved attention mechanisms, and advanced data augmentation techniques. Trained on 4000 sonar images, the model excels in delineating pile structures and effectively identifying underwater targets. Experimental data show that the model can achieve good target identification results in similar experimental scenarios, with a 96.89% accuracy rate for underwater target recognition.

Keywords: AquaPile-YOLO; multi-scale feature fusion; deep learning; sonar image; underwater target recognition; attention mechanism

check for updates

Academic Editor: Jaroslaw Tegowski

Received: 12 December 2024
Revised: 13 January 2025
Accepted: 16 January 2025
Published: 22 January 2025

Citation: Xu, Z.; Wang, R.; Cao, T.; Guo, W.; Shi, B.; Ge, Q. AquaPile-YOLO: Pioneering Underwater Pile Foundation Detection with Forward-Looking Sonar Image Processing. *Remote Sens.* **2025**, *17*, 360. https://doi.org/10.3390/rs17030360

1. Introduction

The detection of underwater pile foundations is important for harbor channel operations and marine engineering [1]. Traditionally, visual examinations by divers have been the main method for identifying underwater pile foundations, but this has limitations including poor safety, high cost, and low efficiency [2,3]. The development of high-resolution sonar imaging technology has opened new possibilities for underwater target detection by offering advantages such as long-range detection capabilities and real-time imaging [4]. However, due to the imaging principles of sonar technology and the impact of underwater environments, sonar images often exhibit high noise, poor contrast, and structural distortions, making the accurate detection and identification of underwater targets difficult [5,6].

The development of underwater pile foundation detection technology has garnered significant attention in the realms of maritime engineering and environmental monitoring. Over the past few decades, underwater target detection using high-resolution sonar imaging has progressed significantly. Early methods focused on feature extraction and enhancement techniques, such as mathematical morphology and level-set methods, to address the inherent noise and resolution issues of sonar imagery. With the advent of deep learning, innovations like the Mask R-CNN and improved YOLO frameworks have emerged, offering enhanced accuracy and robustness. Despite these advancements, key challenges remain, including the detection of small and densely packed targets under varying environmental conditions. This study addresses these challenges by integrating multi-scale feature fusion and attention mechanisms into the AquaPile-YOLO framework. These enhancements are pivotal for the real-time detection and precise identification of underwater pile foundations, enabling significant improvements in sonar image analysis.

Early research on sonar image processing primarily focused on feature extraction and image enhancement [5,7]. For instance, Lu et al. provided a comprehensive review of feature extraction technology for underwater targets using active sonar technology, establishing theoretical foundations for sonar image processing [2]. Subsequently, Calder et al. presented a novel concept for underwater identification of side-scan sonar images—a Bayesian approach to target detection. These early investigations established foundations for understanding and interpreting sonar images [3]. The application of computer vision technologies has enhanced sonar image processing. Foresti et al. proposed an underwater image target recognition method based on a computer vision system, employing computer vision analysis of sonar data [4]. Liu et al. investigated the application of mathematical morphology in acoustic image processing, proving the utility of morphological approaches for image enhancement and edge identification [6].

Deep learning has revolutionized sonar image processing, replacing older methods such as level sets [8], Markov random fields (MRFs) [9], and Curvelet transform [10]. Intelligence in sonar image processing has emerged as the most significant development trend [11]. Intelligence has improved target identification accuracy and efficiency under complex underwater situations [12,13]. Advances in image resolution and quality have made forward-looking sonar broadly applicable in engineering applications [14] like seabed sediment classification [15] and mine target detection [16,17]. Valdenegro-Toro et al. applied convolutional neural networks to target detection and recognition in forward-looking sonar images, initiating deep learning applications in sonar image processing [18]. Zhu et al. addressed the challenge of limited sonar data by proposing a deep network classification algorithm for identifying small bottom targets in high-resolution underwater sonar images, demonstrating the effectiveness of deep learning in small target detection [19].

Most deep learning-based sonar image detection methods rely on sliding window feature extraction, employing various computer vision techniques such as boosted classifiers [20], machine learning classifiers [21–24], and template matching [25,26]. However, these methods often perform poorly outside of the training set, especially in challenging scenarios like underwater tiny target recognition [14–16,27]. Recent research has proposed numerous innovations to address these challenges. For instance, Fan et al. [28] introduced an improved Mask R-CNN method for underwater object detection in forward-looking sonar images, achieving high accuracy. Zhang et al. [29] emphasized the importance of sonar image registration and proposed an improved CNN for learning similarity functions, significantly enhancing model performance. Additionally, Xie et al. [30] released a multibeam forward-looking sonar image dataset, providing a benchmark for target detection. By integrating traditional methods' strengths with deep learning advancements, ongoing research aims to address these challenges, focusing on improving model generalization, efficiency, and robustness for sonar image processing applications.

Building on previous research, Zhang et al. proposed an improved YOLOv5 network for forward-looking sonar images [31], incorporating transfer learning and optimized clustering algorithms. Gaspar et al. have developed unsupervised methods for feature-based place recognition in poor visibility conditions [32], while Jiao et al. proposed the PLUD (Push the right Logit Up and the wrong logit Down) approach to improve sonar image feature representation for open-set and long-tail recognition challenges [33]. Li et al.

introduced TransYOLO, a new forward-looking sonar image target detector based on a TFFN feature fusion network with a transformer stack structure [34].

Leveraging these advancements, this paper proposes an underwater pile foundation detection approach for forward-looking sonar images named AquaPile-YOLO, which is an enhancement of the YOLOv5 algorithm. The AquaPile-YOLO algorithm is designed to overcome the aforementioned challenges by integrating multi-scale feature fusion and attention mechanisms. These enhancements are particularly beneficial for detecting small targets within sonar images. Additionally, the application of data augmentation techniques serves to bolster the model's robustness and generalization capabilities. The training dataset, comprising 4000 sonar images, underwent a series of augmentations including random cropping, rotation, and the introduction of noise to improve the model's adaptability across diverse environmental conditions.

This study proposes AquaPile-YOLO, an advanced algorithm for detecting underwater pile foundations in forward-looking sonar images. By integrating multi-scale feature fusion and attention mechanisms, the proposed method aims to improve detection accuracy and robustness for real-time applications. The ultimate goal is to overcome existing limitations in sonar-based target detection, enabling more reliable and efficient underwater engineering and environmental monitoring applications.

2. Methods

2.1. Forward-Looking Sonar

A forward-looking sonar is an imaging sonar that uses transducers to emit and receive sound waves, forming images from the intensity of sound wave reflections off of underwater targets [26]. Like an optical camera, a forward-looking sonar generates images, but sonar images typically show an overhead view rather than the frontal perspective of an optical camera. Figure 1 illustrates the imaging principle, depicting the 2D reconstruction of a 3D underwater target by a forward-looking sonar.

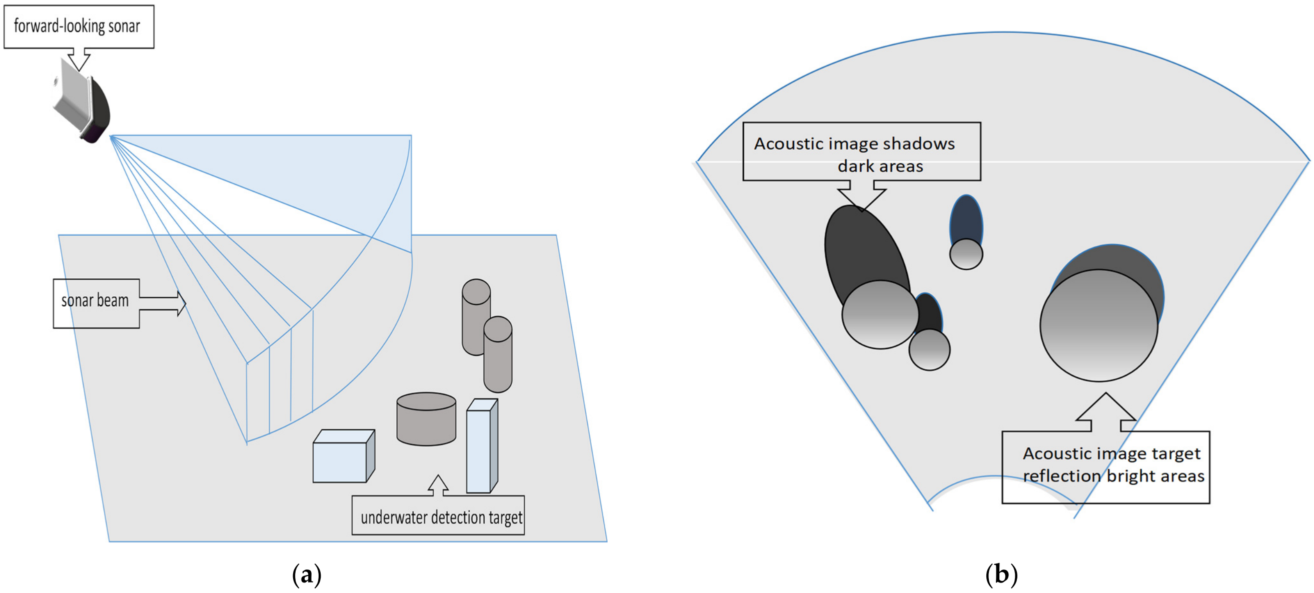

forward-looking sonar
sonar beam
underwater detection target
(a)
Acoustic image shadows dark areas
Acoustic image target reflection bright areas
(b)

Figure 1. A diagram of a 2D reconstruction of an underwater 3D target using a forward-looking sonar: (**a**) A schematic of the FLS operational principle in an underwater environment, depicting the acoustic imaging process; (**b**) An illustration of the 2D reconstruction process, transforming 3D target data into a planar representation as captured by the sonar.

The equipment used in the experiments for this paper is the HY1645 model forward-looking sonar, manufactured by Haiying Marine in Wuxi, China [35]. The sonar utilizes two-dimensional acoustic imaging technology to obtain real-time, high-resolution images of underwater targets (including bearings and distances) for the autonomous recognition and transmission of information. It can meet the needs of autonomous detection in complex, low-visibility, shallow water environments. To meet engineering demands for portability, the device incorporates a novel sparse array design for multibeam imaging sonar systems, reducing the number of transducers while preserving imaging performance. This minimizes the number of transducers in the array while maintaining multibeam imaging performance [36]. A schematic diagram of the fan-scan function for detecting underwater pile foundation targets by a forward-looking sonar is shown in Figure 2.

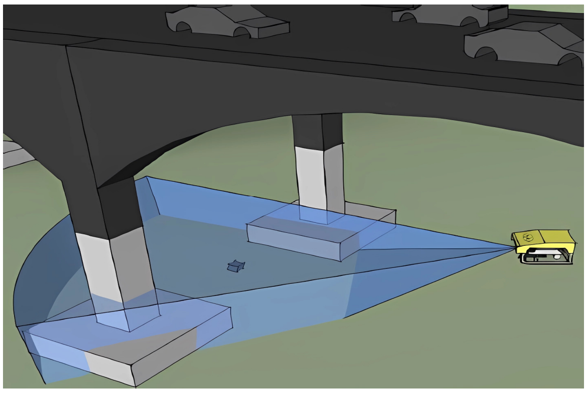
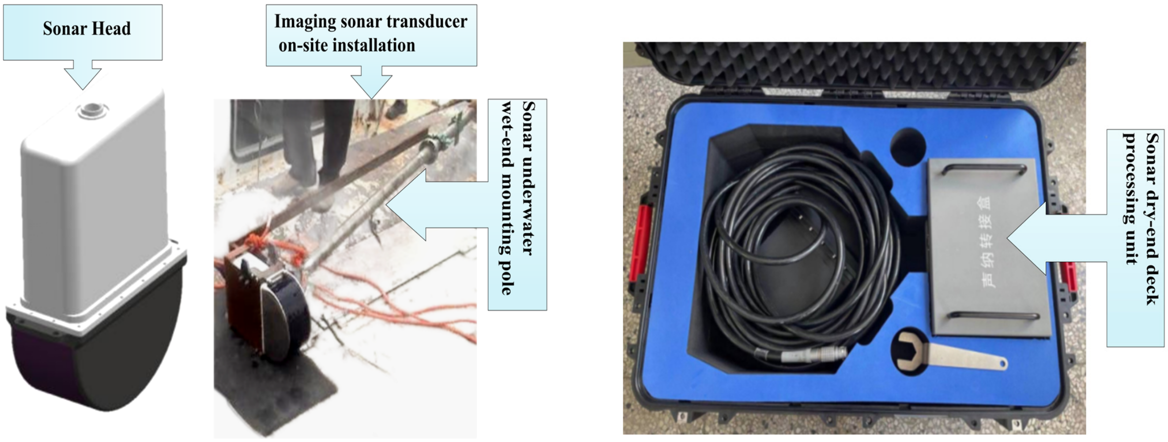


Figure 2. A schematic diagram of the underwater detection capabilities of a forward-looking sonar.

The system primarily consists of an underwater transducer, a transmitter/receiver module, a data acquisition processor, and acquisition software, among other components. Figure 3 illustrates a photo of the HY1645 imaging sonar transducer and its on-site installation. In the photo, the black part of the transducer is responsible for the reception and transmission of underwater acoustic signals, while the white part encloses the receiver and transmitter modules along with their associated circuitry. The entire assembly is encapsulated in waterproof housing for integrated packaging and communicates and receives power from the exterior through a single cable.


Sonar Head
Imaging sonar transducer on-site installation
Sonar underwater wet-end mounting pole
Sonar dry-end deck processing unit

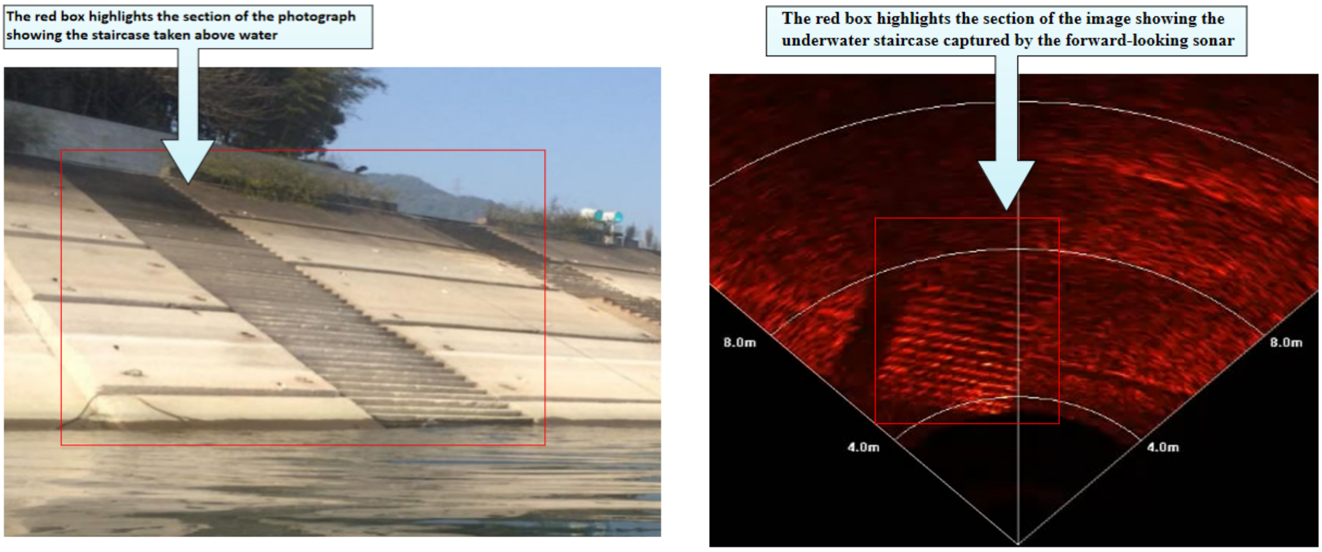
(**a**) (**b**)

Figure 3. The composition of the HY1645 forward-looking sonar system: (**a**) The wet end of the sonar, designed for underwater acoustic signal emission and reception; (**b**) The dry end components of the HY1645 sonar, including the data processing unit and associated cabling.

The main technical parameters and performance indicators of the HY1645 forward-looking sonar are presented in Table 1. A significant characteristic of forward-looking sonars is that, in most cases, the distance and bearing of objects can be directly read from the sonar data, but the elevation of underwater targets is lost. As a result, the image information from forward-looking sonars is typically challenging to interpret. For instance, during the detection of an underwater stepped structure at a hydropower station using the HY1645 forward-looking sonar, Figure 4a shows the surface photo of the stepped structure, while Figure 4b displays the corresponding underwater sonar data collected by the two-dimensional imaging sonar.

Table 1. Technical specifications of HY1645 forward-looking sonar.

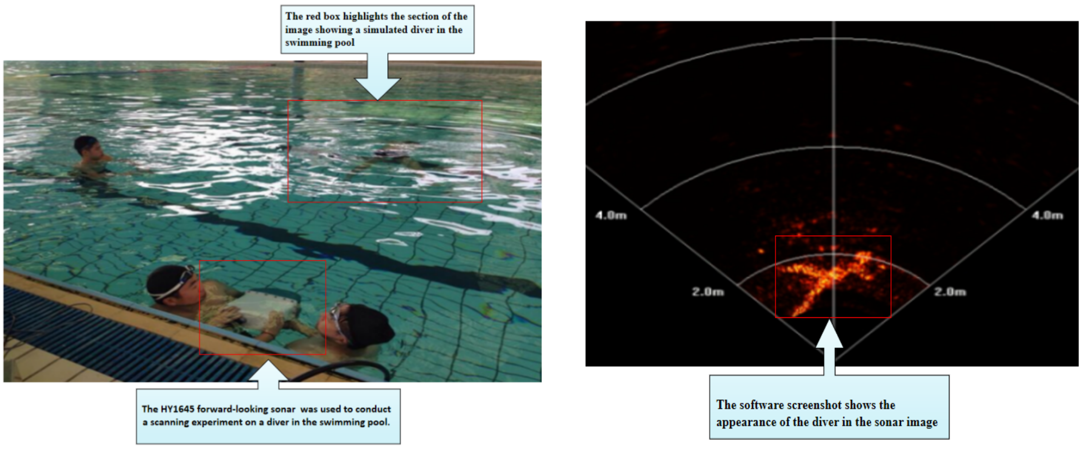
Parameter	Value
Operating Frequency	450 kHz
Field of View	90° × 20°
Maximum Range	100 m
Beam Width (Horizontal × Vertical)	1° × 20°
Number of Beams	538
Beam Spacing	0.17°
Range Resolution	2.5 cm
Maximum Sampling Rate	15 Hz


The red box highlights the section of the photograph showing the staircase taken above water
The red box highlights the section of the image showing the underwater staircase captured by the forward-looking sonar
8.0m
8.0m
4.0m
4.0m

(**a**) (**b**)

Figure 4. Sonar scanning experiment of the underwater ladder structure: (**a**) Photos of the above-water part of the ladder underwater structure sonar scanning experiment; (**b**) Forward-looking underwater ladder scanning experimental sonar data.

Acoustic imaging cannot capture the true color of detected objects, yielding purely grayscale initial data. Yellow sonar images are pseudo-colored, enhanced in contrast via software processing. The HY1645 imaging sonar can scan both static and dynamic underwater targets, like divers. Figure 5 illustrates the use of an imaging sonar to simulate the monitoring of a diver in a swimming pool.


The red box highlights the section of the image showing a simulated diver in the swimming pool

The HY1645 forward-looking sonar was used to conduct a scanning experiment on a diver in the swimming pool.

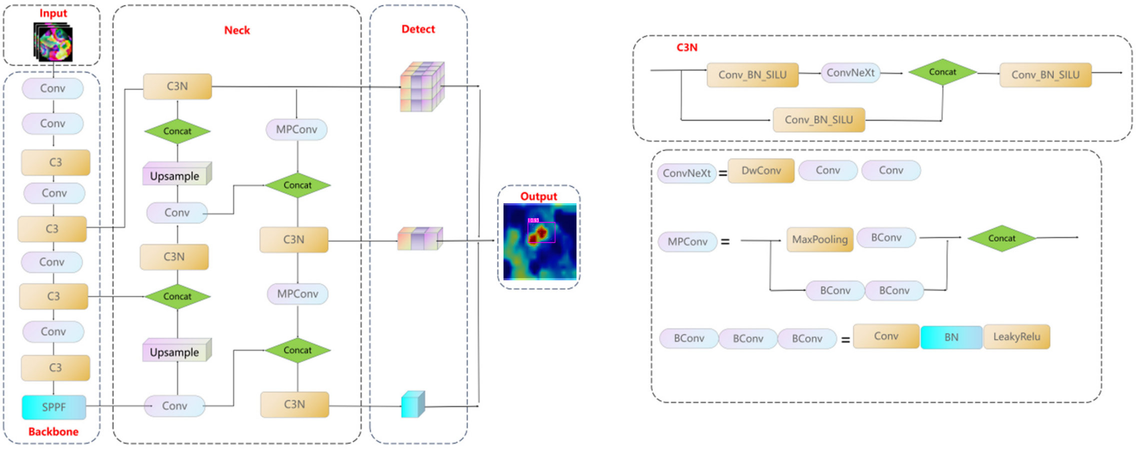
The software screenshot shows the appearance of the diver in the sonar image

(**a**) (**b**)

Figure 5. Sonar scanning experiment of the diver's pool: (**a**) The forward-looking sonar scanning diver swimming pool experiment; (**b**) The forward-looking sonar scans the sonar image data of the diver's pool experiment.

2.2. *AquaPile-YOLO Network*

The YOLO (You Only Look Once) network [37] is a revolutionary real-time object detection system that can predict the positions and categories of objects in an image through a single forward propagation. YOLOv5 is an efficient object detection algorithm renowned for its speed and superior performance. Figure 6 illustrates the structure of the AquaPile-YOLO network. In recent years, through continuous updates and iterations [38,39], the YOLO network has been widely applied in engineering projects due to its stability. However, forward-looking sonar images present unique challenges, requiring adaptations for effective detection. This study aims to enhance the performance of AquaPile-YOLO in underwater pile foundation detection tasks by introducing a series of innovative improvements. These enhancements were designed to adapt to the particularities of forward-looking sonar images and increase the detection accuracy of underwater pile foundation targets.



(**a**) (**b**)

Figure 6. The AquaPile-YOLO network structure pipeline diagram: (**a**) This panel presents an overview of the AquaPile-YOLO's architecture, illustrating the comprehensive workflow from input to output, and highlighting the integration of multi-scale feature fusion, attention mechanisms, and other key components that facilitate the detection of underwater pile foundations; (**b**) This panel zooms in on specific modules within the AquaPile-YOLO network, detailing the internal structure and connectivity of the components, such as the C3 Module with CBAM Attention, MPConv Module, and C3N Module, which are crucial for enhancing the network's performance in processing forward-looking sonar images.

2.2.1. Data Augmentation

To enhance the model's generalization and robustness, data augmentation techniques were employed. Operations such as rotation, scaling, flipping, and adding noise to the training images simulate the complexity of underwater environments, effectively increasing the diversity of the training data.

It was assumed that the sonar image dataset is divided into groups of four sub-images I1, I2, I3, and I4, each of a size of H × W. Through random cropping and flipping operations, each sub-image can generate a new sub-image I1′, I2′, I3′, and I4′. These sub-images were then concatenated into a larger image Imosaic of size 4H × 4W. The concatenation operation can be expressed as follows:

$$I_{\text{mosaic}}(x,y) = \begin{cases} I'_1(x-2H, y-2W) \text{ if } x \in [0,2H) \text{ and } y \in [0,2W) \\ I'_2(x-H, y-2W) \text{ if } x \in [2H,3H) \text{ and } y \in [0,2W) \\ I'_3(x-2H, y-H) \text{ if } x \in [0,2H) \text{ and } y \in [2W,3W) \\ I'_4(x-H, y-H) \text{ if } x \in [2H,3H) \text{ and } y \in [2W,3W) \end{cases} \quad (1)$$

where (x,y) represents the coordinate position in Imosaic.

2.2.2. Transfer Learning Strategy

Considering the scarcity of sonar image data, a transfer learning strategy was adopted. A pre-trained AquaPile-YOLO model, initially trained on a large dataset like ImageNet, served as the starting point. Then, by fine-tuning AquaPile-YOLO on the limited sonar image data, the model's performance was quickly enhanced. The transfer learning strategy by Huo et al. [40] for side-scan sonar image classification and target recognition is referenced.

Let the source domain be D_s = {χ,P(X)} and the target domain be D_t = {χ′,P(X′)}, where χ and χ′ represent the feature spaces, and P(X) and P(X′) represent the marginal probabilities. The task T is defined by the label space y and the target prediction function f(x). The goal of transfer learning is to improve the performance of the prediction function f_t for the target task T_t by discovering and transferring knowledge from D_s and T_s.

During the pre-training phase, a deep network F was trained on the source domain to learn general feature representations.

$$F^* = \text{argmin}_F L(F(X^s), Y^s) \quad (2)$$

where L is the loss function, X^s and Y^s are the input and label of the source domain, respectively, and F^* is the pre-trained network.

In the transfer phase, the pre-trained network F^* was transferred to the target domain and adapted to the target task through fine-tuning.

$$F' = \arg\min_F L\big(F(X^t), Y^t\big) \quad (3)$$

where X^t and Y^t are the inputs and labels for the target domain, respectively.

2.2.3. Multi-Scale Feature Fusion

Multi-scale feature fusion techniques were introduced into the AquaPile-YOLO to address the variability in target sizes within forward-looking sonar images. This strategy enhanced the model's ability to recognize targets of various scales by integrating feature maps at different resolutions. A Feature Pyramid Network (FPN) structure was employed to effectively combine deep semantic information with shallow detail information, thereby improving the detection accuracy of small targets. The feature fusion can be expressed as follows:

$$\mathrm{F_{fuse}} = F_{\mathrm{upsample}}(\mathrm{F_d}) \oplus F_{\mathrm{downsample}}(\mathrm{F_c}) \quad (4)$$

where $\mathrm{F_d}$ represents the deep-layer feature map and $\mathrm{F_c}$ represents the shallow-layer feature map. $\mathrm{F_{upsample}}$ and $\mathrm{F_{downsample}}$ denote the upsampling and downsampling operations, respectively, while the symbol $\oplus$ signifies the operation of feature fusion.

The upper-level feature maps contain stronger semantic information due to the deeper network layers, while the lower-level features suffer less loss of positional information due to fewer convolutional layers. The FPN structure performs top–down upsampling to ensure that the bottom-level feature maps contain stronger semantic information (the backbone in Figure 6). Conversely, the PAN (Path Aggregation Network) structure performs bottom–up downsampling, enabling the top-level features to retain positional information (neck in Figure 6). The fusion of these two features ensures that feature maps of different scales contain both semantic and spatial information, thereby facilitating accurate predictions for images of various sizes.

Fine-tuning further trained the network on the target domain data, adjusting the network parameters to improve performance.

$$\theta' = \theta - \eta \nabla_{\theta} \mathrm{L}\left(\mathrm{F}\left(\mathrm{X^t};\theta\right),\mathrm{Y^t}\right) \quad (5)$$

where θ is the network parameter, η is the learning rate, and ∇ stands for the gradient.

2.2.4. Attention Mechanism

The AquaPile-YOLO network incorporates advanced attention mechanisms to enhance the model's ability to focus on salient regions within the image, particularly in complex underwater environments characterized by noise and occlusions. This was achieved through the integration of the Convolutional Block Attention Module (CBAM) into the YOLOv5 network architecture. In this section, we will discuss the role of the C3 module (CSP Bottleneck with 3 convolutions) [41], MPConv, and the C3N module in enhancing the attention mechanisms of the AquaPile-YOLO network.

(1) C3 Module with CBAM Attention

An attention mechanism called CBAM was incorporated into the AquaPile-YOLO network to enhance the model's focus on key areas within the image. This mechanism comprises spatial and channel attention modules that adaptively adjust the weights of the feature maps, enhancing the model's response to target areas, especially in complex underwater environments with noise and occlusions.

For example, the channel attention for an input feature map Fattn is given by the following:

$$\mathrm{F_{attn}} = \sum_{\mathrm{c}} \mathrm{A_c} \cdot \mathrm{F_c} \quad (6)$$

where $\mathrm{A_c}$ is the attention weight of the c channel, typically calculated using learnable parameters and an activation function σ as follows:

$$\mathrm{A_c}{=}\ \sigma(\mathrm{W}{\cdot}\mathrm{F_c} + \mathrm{b}) \quad (7)$$

where W and b are the weight parameters and bias parameters in the deep network, respectively.

The C3 module comprises a main branch (primary path) and a shortcut branch (skip connection), which are merged at the output [41]. The main branch typically includes multiple Bottleneck layers, sequentially stacked to increase the network's depth and representational capacity. By replacing the default Bottleneck layers in the C3 module with

CBAM modules and iteratively creating multiple CBAM Bottleneck layers, the integration of CBAM attention mechanisms within the C3 module was achieved (C3-CBAM in Figure 6). The C3-CBAM module retains the advantages of the C3 module, such as efficient feature extraction and partial gradient flow sharing, while significantly enhancing feature representation through the CBAM's channel and spatial attention mechanisms. This synergistic combination endowed YOLOv5 with higher accuracy and robustness in object detection tasks, thereby improving the overall model performance on sonar objects.

(2) MPConv Module

The MPConv (Multi-Path Convolution) module is a novel architectural component introduced in the AquaPile-YOLO network to address the challenges posed by the diverse scales and orientations of underwater targets in sonar images. MPConv is designed to capture a rich set of features by processing the input data through multiple parallel convolutional paths with different kernel sizes and aspect ratios [42]. Each path is tailored to capture specific spatial hierarchies, allowing the network to represent a wide range of underwater structures effectively. The outputs from these parallel paths are then concatenated, forming a comprehensive feature representation that encapsulates both local and global contextual information. This multi-path processing approach enabled the AquaPile-YOLO network to achieve superior performance in detecting targets of varying sizes and complexities within sonar imagery.

(3) C3N Module

The C3N module, building on the strengths of the C3 module, introduces an innovative structure combining depth-separable convolution with a novel inverted Bottleneck design, inspired by the ConvNeXt architecture [43,44]. The C3N module consists of three convolutional layers followed by multiple ConvNeXt blocks, enabling efficient parameter utilization and enhanced feature correlation capture while mitigating information loss during dimensionality compression. The inverted Bottleneck structure of the C3N module, with a wider central section and narrower endpoint, empowers effective feature correlation capture and efficient feature space transformation processing. This results in robust feature extraction capability, particularly beneficial for detecting small, densely packed targets in sonar images, despite the imaging limitations of sonar technology.

By integrating these advanced modules—C3 with CBAM, MPConv, and C3N—the AquaPile-YOLO network achieved a heightened level of attention and discrimination, enabling it to excel in the detection of underwater pile foundation targets within forward-looking sonar images [45].

2.2.5. Loss Function Optimization

The loss function plays a crucial role in object detection tasks. The loss function for AquaPile-YOLO was optimized based on the characteristics of sonar image targets, employing a composite loss function that guides model training more comprehensively through classification loss L_{cls}, regression loss L_{reg}, and objectness loss L_{obj}.

The total loss function is given by the following:

$$\mathrm{L_{sonar}} = \frac{1}{\mathrm{N_{pos}}}\left(\mathrm{L_{cls}} + \mathrm{L_{reg}} + \mathrm{L_{obj}}\right) \tag{8}$$

where $\mathrm{N_{pos}}$ is the number of positive samples, $\mathrm{I}\{\cdot\}$ is the indicator function, $\mathrm{L_{Focal}}$ is the focal loss for classification, $\mathrm{L_{IoU}}$ is the IoU loss for regression, and $\mathrm{L_{BCE}}$ is the binary cross-entropy loss for objectness.

2.2.6. Soft-NMS (Soft-Non-Maximum Suppression)

In this study, the Soft-NMS algorithm [46] was used to improve the object detection process of AquaPile-YOLO. Soft-NMS adjusts the scores of detection boxes using a Gaussian function for continuous decay instead of simply setting the scores of overlapping detection boxes to zero, thereby improving the accuracy and robustness of small target detection.

In traditional NMS, given a set of detection boxes B = {b_1,. . .,b_N} and corresponding scores S = {s_1,. . .,s_N}, the algorithm first selects the box with M the highest score, then removes all other boxes with an overlap higher than the threshold of N_t with M. This process is then recursively applied to the remaining boxes. Soft-NMS proposes a different approach by adjusting the score s_i of the detection box b_i using the following Formula (9):

$$s_i' = s_i \cdot e^{-\mathrm{IOU}(M,b_i)^2/\sigma} \forall b_i \notin D \quad (9)$$

where IOU(M,b_i) represents the Intersection over Union between the detection boxes M and b_i, and σ is a parameter controlling the speed of score decay.

3. Experiments

3.1. Experimental Design

The purpose of this experiment was to validate the effectiveness of the proposed AquaPile-YOLO algorithm for underwater pile foundation detection using HY1645 forward-looking sonar images. The experimental environment was a designated section of a lake field test site, characterized by water depths ranging from 2 m to 20 m and a substrate primarily composed of sand and gravel, providing a controlled yet representative setting for underwater sonar testing.

The HY1645 forward-looking sonar was installed on a vessel using a lateral straddle mount, as shown in Figure 7, which illustrates the field experiment vessel with the sonar installed. Two devices were fixed onto an installation pole. Due to the weight of the equipment, the structure was designed to grip the vessel's edge from both sides beneath the bow. The installation pole was fixed to the side of the vessel, with the detection sonar located approximately 0.5 m below the water surface.

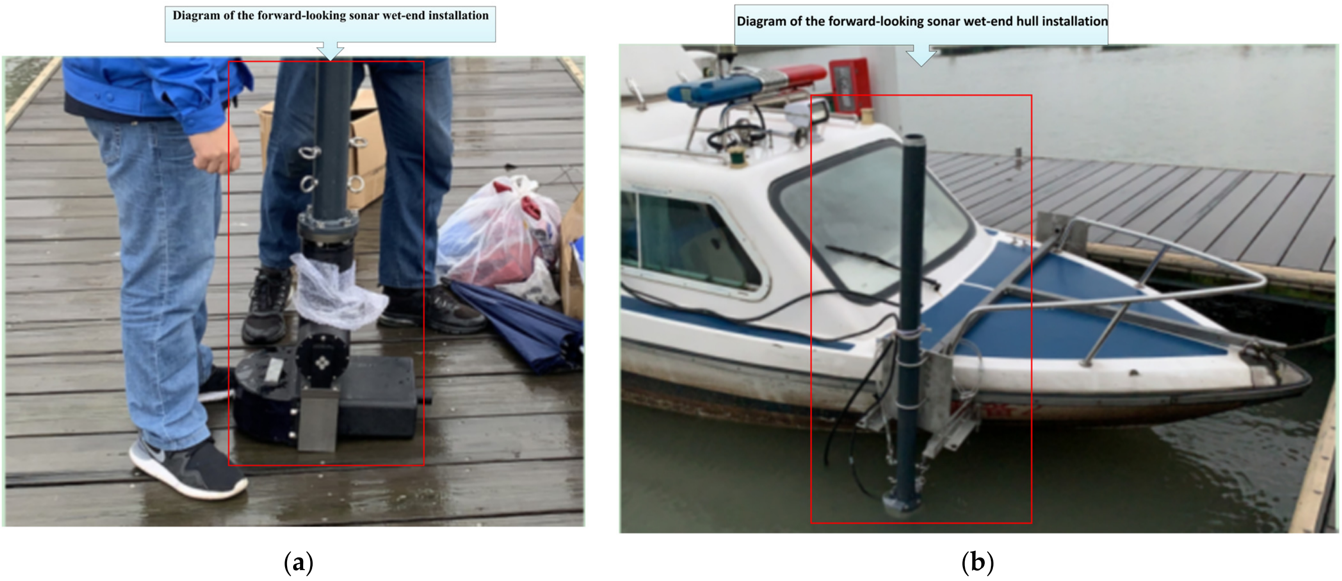
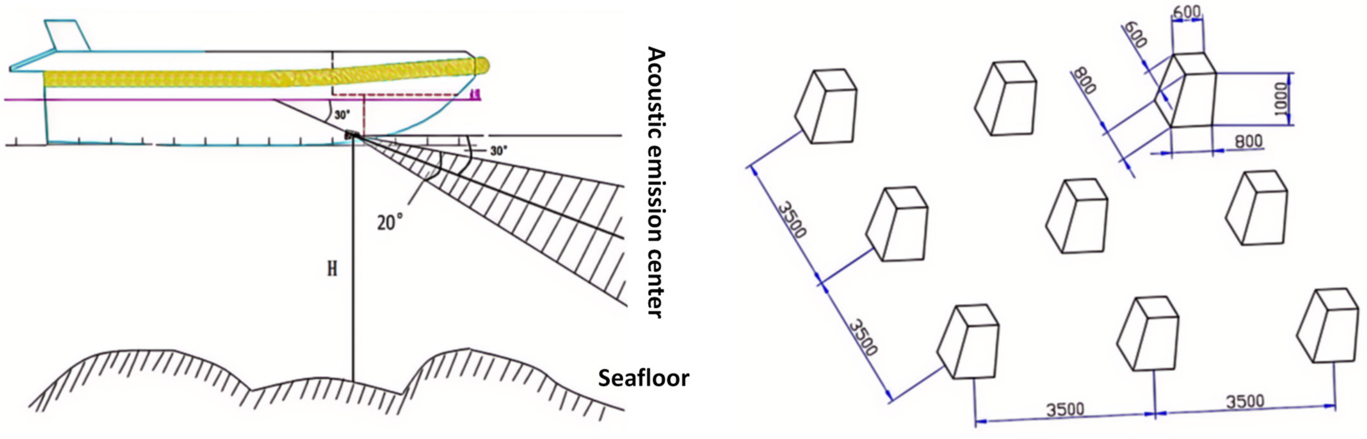


Diagram of the forward-looking sonar wet-end installation
Diagram of the forward-looking sonar wet-end hull installation
(a)
(b)

Figure 7. HY1645 forward-looking sonar field test installation; Forward-looking sonar installation angle (**a**) and target (underwater pile foundation) distribution (**b**) diagram.

To prevent interference from the side lobes of the forward-looking sonar touching the water surface and causing noise, the emission direction of the detection sonar was set to a 30° downward tilt from the water surface, based on the scanning direction of the sonar beam opening angle. The sonar installation angle (left) and the distribution of the target (underwater pile foundation) (right) are shown in Figure 8.



(**a**) (**b**)

Figure 8. Forward-looking sonar installation angle (**a**) and target (underwater pile foundation) distribution (**b**) diagram.

3.2. Data Collection

During the data collection phase, we conducted field experiments at one lake in Wuxi City from November 26th to 27th, 2020. The trials involved multi-distance, multi-directional sonar detection of pre-set targets (track racks) at varying speeds (13 km/h, 15 km/h, and 18 km/h, equivalent to approximately 7, 8, and 10 knots, respectively). Utilizing a gimbaled mount, the sonar was adjusted to an optimal operational attitude to ensure the acquisition of target sonograms in real-time. Data were recorded in the AVI video format and were saved as JPEG/PNG/BMP snapshots for subsequent analysis and algorithm validation.

Data annotation was performed for underwater pile foundation targets in sonar images by conducting continuous long-term detection and comparing them with human observations and mapping charts. The dataset includes two category labels, ”l” and ”r”, using the YOLO format. To construct the training dataset, 4000 sonar images were collected in the field experiment, covering various underwater environments and target conditions. The image data were preprocessed, including grayscaling, noise removal, and contrast enhancement, to improve the accuracy of subsequent target detection.

Due to the scarcity of sonar data, scholars in the field of sonar images have mostly used simulated datasets as the sample space, while actual measured datasets barely exceeded a few hundred images. This paper collected 4000 sonar images on-site as the dataset for deep learning training, which to some extent compensates for the lack of data in previous research in this field.

The original data collected by the HY1645 imaging sonar were in a custom format of acoustic signal data, with “.hca” and “.son” being the two formats. The HAICA.EXE executable program provided by the system is required to read them. The original acoustic signals were transformed into image data in the “.bmp” format. Using the original acoustic data collected by the HY1645 in the field experiment, 4000 two-dimensional sonar images with a pixel resolution of 848 × 600 were generated.

3.3. Experimental Procedure

This study employs a comprehensive dataset, containing a total of 4000 field-measured, forward-looking sonar data images, which were meticulously preprocessed to augment model detection capabilities. In the experiment, the dataset was divided into a training set (3000 images) and a validation set (1000 images). The experimental steps incorporate several key stages: data augmentation, model training, performance evaluation, and systematic recording of results. The data augmentation phase plays a crucial role in enhancing model adaptability across diverse environments, achieved through an array of techniques such as random cropping, rotation, and the strategic introduction of noise. The model training phase was executed within a strictly defined, controlled environment, where parameters including the learning rate and batch size were rigorously monitored.

The experiment was conducted on a system equipped with a high-performance CPU and GPU to ensure efficient operation. The CPU is Intel(R) Xeon(R) Gold 6130, and the GPU is Tesla V100-PCIE-32GB, with 32 GB of video memory. The operating system used is Ubuntu 18.04.5 LTS, and the deep learning framework is torch-2.0.0, as shown in Table 2 for the detailed experimental environment configuration.

Table 2. Experimental environment configuration.

Parameter	Setup
Ubuntu	18.04.5 LTS
Pytorch	2.0.0
Python	3.8
CUDA	11.8
GPU	Tesla V100-PCIE-32GB
CPU	Intel(R) Xeon(R) Gold 6130

In order to enhance the persuasiveness of the experiments, this study conducted a series of parameter adjustments based on the AquaPile-YOLO model and performed multiple experimental tests, ultimately selecting the hyperparameter settings as shown in Table 3.

Table 3. Hyperparameters during training.

Parameter	Setup
Epoch	300
Batch	32
NMS IoU	0.6
Initial Learning Rate	0.01
Final Learning Rate	0.01
Momentum	0.937
Weight Decay	0.0005

The formulas are as follows. Regular evaluations were undertaken using a validation set to ensure the model's performance was accurately gauged. Key metrics like precision, recall, and mAP were systematically recorded during this stage. The formulas are as follows.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$AP = \int_0^1 P(R)dR \tag{12}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{n} AP_i \quad (13)$$

where TP is the number of correctly predicted positive samples, FP is the number of negative samples incorrectly predicted as positive, and FN is the number of positive samples incorrectly predicted as negative. Moreover, average precision (AP) is the calculation of the area under the accuracy–response rate curve for a certain category. mAP is an auxiliary to the AP of all categories and can be used to evaluate the model's detection performance for all categories. In Formula (13), n is the number of categories; AP(j) is the AP of the jth category.

To ensure methodological rigor and reproducibility, all experimental settings and parameters were painstakingly documented. Furthermore, the entire experiment was repeated multiple times in order to confirm the consistency and reliability of the results obtained. Potential biases and errors that could arise during the course of the study were identified and discussed, along with the corresponding mitigation strategies proposed. This thorough experimental procedure aimed to provide a transparent and replicable guide for scholars seeking to replicate the study's setup, as well as to harness the enhanced capabilities of the AquaPile-YOLO model within their own research endeavors.

4. Results

4.1. Ablation Studies

In order to analyze the influence of different improvement strategies on the performance of model detection, three groups of experiments were designed to complete the training and testing under the premise of ensuring the same data set and training parameters and the experimental results are shown in Table 4.

Table 4. Results of ablation experiments.

Multi-Scale Feature Fusion	CBAM	Sonar Loss	Soft-NMS	Precision	Recall	mAP50	mAP50-95
×	×	×	×	0.886	0.76	0.789	0.517
✓	×	×	×	0.9	0.764	0.8	0.521
×	✓	×	×	0.912	0.764	0.808	0.524
✓	✓	×	×	0.919	0.771	0.811	0.525
✓	✓	✓	×	0.896	0.785	0.819	0.528
✓	✓	✓	✓	0.888	0.798	0.821	0.529

When only CBAM was enabled, the precision further increased to 0.912, while the recall remained at 0.764. The mAP50 improved to 0.808, and the mAP50-95 increased to 0.524. This demonstrates the significant effect of the CBAM on enhancing model performance. By combining MSFF and the CBAM, the performance continued to improve, with precision reaching 0.919, recall increasing to 0.771, mAP50 rising to 0.811, and mAP50-95 reaching 0.525. This combination clearly outperforms the use of MSFF or CBAM alone.

After introducing Sonar Loss, the precision slightly decreased to 0.896, but the recall improved to 0.785. The mAP50 increased to 0.819, and the mAP50-95 reached 0.528. This indicates that Sonar Loss is helpful in improving the recall rate and overall performance of the model, although it may slightly impact accuracy.

Finally, with all improvements (including Soft-NMS) enabled, the precision was 0.888, recall increased to 0.798, mAP50 reached 0.821, and mAP50-95 also increased to 0.529. Despite a slight decrease in accuracy, the improvements in recall and mAP values reflect the enhancement of overall detection performance.

In conclusion, by combining techniques such as multi-scale feature fusion, CBAM, Sonar Loss, and Soft-NMS, AquaPile-YOLO achieved improvements in various performance metrics, particularly in mAP, the most convincing indicators. These enhancements effectively enhance the detection capabilities of YOLOv5.

4.2. Comparisons

After ablation studies, the AquaPile-YOLO model was tested by comparative experiments. The test results showed that the model achieved an identification accuracy rate of 96.89% for underwater targets, confirming the effectiveness and reliability of the proposed method in actual underwater pile foundation detection.

This experiment compared the performance of five object detection algorithms, SSD300, YOLOv3, Faster R-CNN, Cascade R-CNN, and AquaPile-YOLO, on sonar images. The test results for each algorithm are shown in Table 5. Additionally, we compared our results with the recently published Underwater Acoustic Target Detection (UATD) dataset. This dataset includes identification results for underwater objects such as a ball, cube, tire, sc (square cage), and cc (circle cage). As shown in Table 6, the AquaPile-YOLO model performed superiorly across these categories, further validating its efficacy in various underwater detection scenarios.

Table 5. Comparison of AquaPile-YOLO with other models.

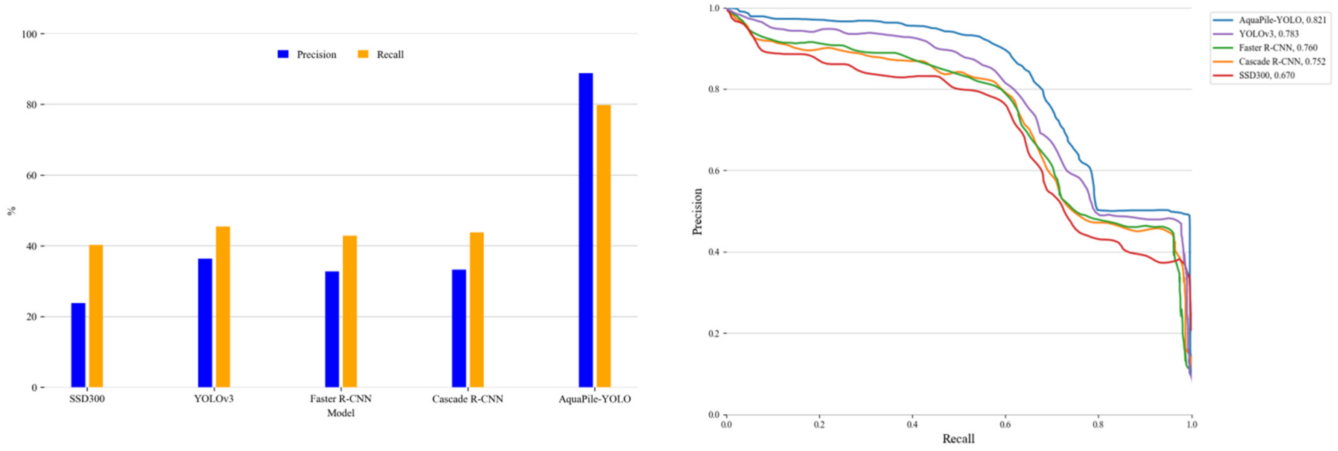
Model	Precision	Recall	mAP@50	Params/M	FPS
SSD300	0.238	0.403	0.670	23.88	9.1
YOLOv3	0.364	0.455	0.783	61.52	46.7
Faster R-CNN	0.328	0.429	0.760	41.35	19.4
Cascade R-CNN	0.333	0.438	0.752	69.15	15.5
AquaPile-YOLO	0.888	0.798	0.821	46.60	111.1

Table 6. Detection results of underwater targets with different scenarios.

Model	AP (Ball)	AP (Cube)	AP (Tyre)	AP (sc)	AP (cc)	AP (Pile)
Faster-RCNN (Resnet-18)	0.869	0.717	0.847	0.547	0.666	-
Faster-RCNN(Resnet-50)	0.870	0.686	0.889	0.621	0.538	0.328
Faster-RCNN(Resnet-101)	0.865	0.697	0.840	0.572	0.491	0.333
YOLOv3 (Darknet-53)	0.860	0.669	0.874	0.470	0.519	-
YOLOv3 (MobilenetV2)	0.868	0.573	0.738	0.518	0.498	0.364
AquaPile-YOLO	-	-	-	-	-	0.888

As shown in Figure 9a, the comparative analysis indicates that AquaPile-YOLO outperforms other state-of-the-art object detection models, including YOLOv3, Faster R-CNN, Cascade R-CNN, and SSD300, in terms of both precision and recall. Precision, which quantifies the proportion of true positive detections among all detected samples, and recall, which measures the model's ability to detect all actual positive instances, are critical metrics for object detection systems. AquaPile-YOLO achieves a precision of 0.888 and a recall of 0.798, with a mAP@50 score of 0.821, indicating its exceptional ability to identify underwater pile foundations while minimizing false positives accurately. This high level of precision and recall suggests that AquaPile-YOLO is particularly robust in scenarios requiring reliable underwater detection.

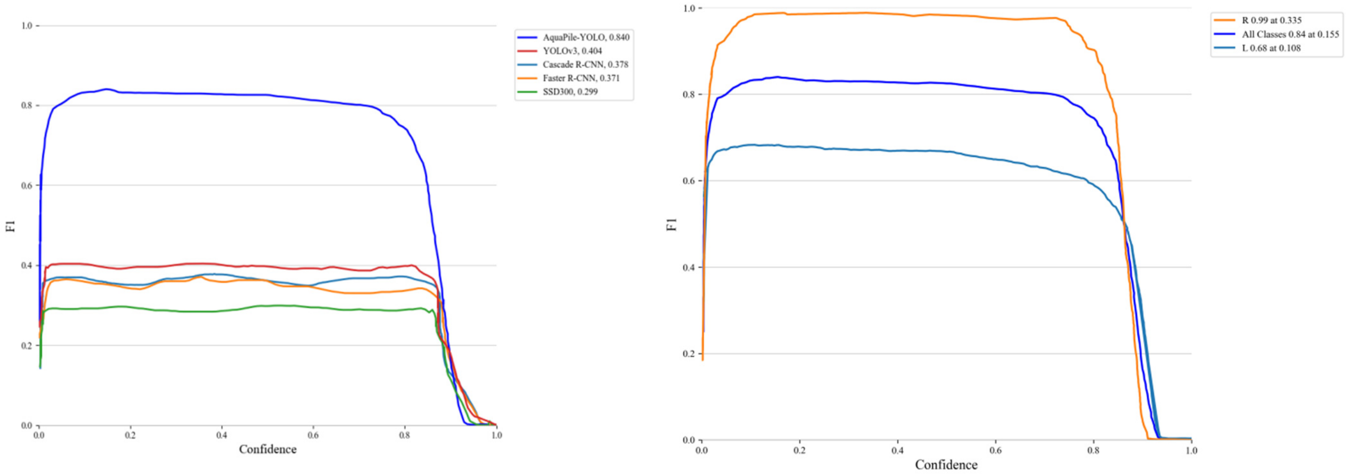
Figure 9b provides a detailed comparison of recall performance among the same set of object detection models, further emphasizing AquaPile-YOLO's superiority. With a recall value of 0.821, AquaPile-YOLO demonstrates its effectiveness in detecting all instances of underwater targets, outperforming YOLOv3 (0.783), Faster R-CNN (0.760), Cascade R-CNN (0.752), and SSD300 (0.670). This superior recall performance indicates that AquaPile-YOLO is more reliable in identifying underwater targets, making it highly suitable for applications where high recall is essential for operational success. The high recall rate is particularly crucial in underwater environments, where missing a target could have significant consequences, thus highlighting AquaPile-YOLO as the preferred model for critical detection tasks.



(**a**) (**b**)

Figure 9. (**a**) Bar chart comparison of precision–recall for different models; (**b**) Comparison of object detection models.

Figure 10a presents a compelling comparison of F1 performance for pile foundation detection using forward-looking sonar images across various algorithms. The F1 score, a balanced metric harmonizing both precision and recall, is depicted at varying confidence thresholds. This composite score provides a comprehensive measure of a model's exactness and completeness in detection. AquaPile-YOLO exhibits notably high F1 scores, signifying its ability to balance precision and recall. Notably, at a confidence threshold of 0.155, AquaPile-YOLO achieves an F1 score of 0.84, indicating robustness in accurately detecting pile foundations.



(**a**) (**b**)

Figure 10. (**a**) F1 performance comparison of different algorithms for pile foundation detection by forward-looking sonar images; (**b**) F1 performance curve for AquaPile-YOLO.

Figure 10b depicts the F1 performance curve for the AquaPile-YOLO model, illustrating how the model's F1 score fluctuates at different confidence thresholds. The curve represents the interplay between precision and recall, with each point reflecting the precision at various levels of recall. This visualization is instrumental in assessing the model's performance across the entire spectrum of detection confidence. The curve underscores

AquaPile-YOLO's consistently high performance, even at lower confidence thresholds, thereby validating its reliability and effectiveness in real-world applications. The "All classes" average F1 score encapsulates the model's overall efficacy in detecting a diverse range of underwater targets, further solidifying AquaPile-YOLO as a superior choice for sonar-based object detection tasks.

Figure 11 is a heatmap comparison, demonstrating the comprehensive performance of different algorithms on the target detection task. AquaPile-YOLO achieved a high score of 0.93 on this indicator. A comparison of the original image and detection results for each algorithm's heatmap indicates the model's strong comprehensive performance for detecting underwater pile foundation targets under various scenarios. Simultaneously, it shows the model performs well in sonar image target detection tasks, meeting real-time detection speed requirements and significantly improving accuracy. These results support the model in this paper as the preferred algorithm for sonar image target detection.

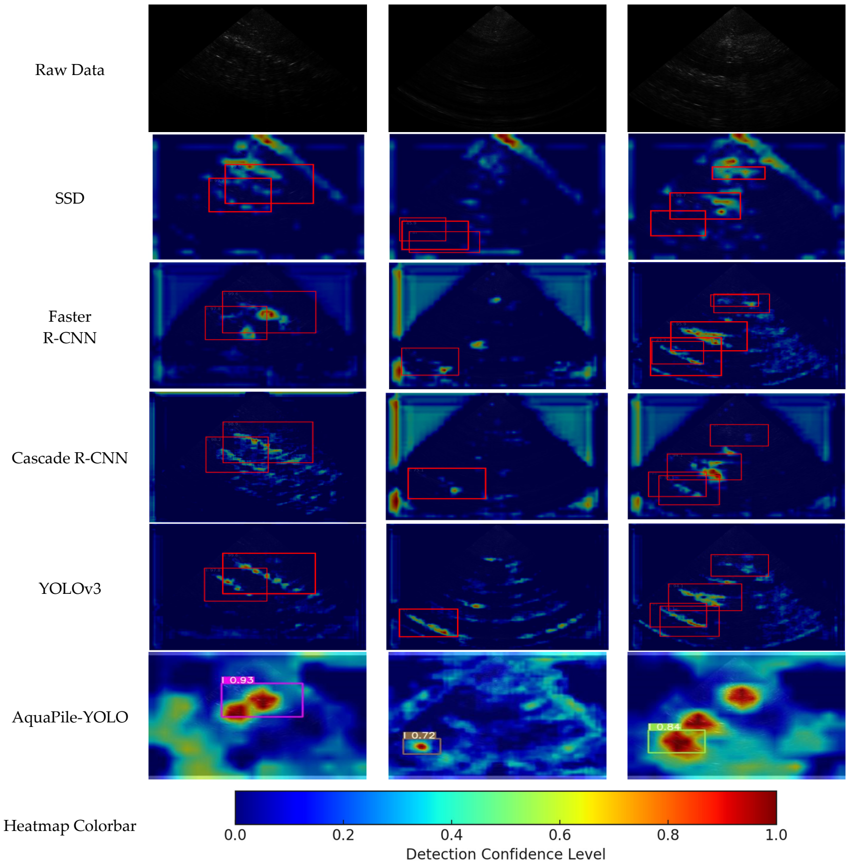


Figure 11. A heatmap comparison of different algorithms for pile foundation detection by forward-looking sonar images. The heatmap illustrates the performance comparison of various detection algorithms, with the red box highlighting the area of interest where the pile foundation targets are detected. Within this box, the intensity of the color indicates the confidence level of the detection, with warmer tones (reds and yellows) signifying higher confidence in the presence of a target.

5. Discussion

This study introduces AquaPile-YOLO, an advanced underwater pile foundation detection method utilizing forward-looking sonar imagery. The proposed method offers several advantages, including significantly improved detection accuracy achieved by the AquaPile-YOLO algorithm. The algorithm effectively captures underwater targets of varying sizes and enhances the detection of small targets, representing a critical advancement in the field.

The principal contributions of this study comprise the following: (1) the development and proposal of the AquaPile-YOLO algorithm, an innovative method for underwater pile foundation detection, which builds upon the foundational architecture of YOLOv5 and incorporates multi-scale feature fusion and attention mechanisms to achieve significantly improved detection accuracy; (2) the application of data augmentation techniques to improve model generalization and robustness; (3) the collation and use of 4000 sonar images as a training dataset, offering plentiful data for model training and validation; and (4) experimental results underscoring the considerable practical application value in detecting underwater pile foundation targets within sonar images.

Specifically, this study addresses the critical challenge of real-time, fast, and accurate template recognition and the detection of underwater pile foundations in sonar images. Key innovations include the following:

- Multi-scale Feature Fusion: By incorporating a multi-scale feature fusion scheme, this study effectively captures underwater targets of varying sizes, thereby improving small target detection accuracy.
- Enhanced Attention Mechanism: The attention mechanism is improved by combining Normalized Weighted Distance (NWD) and Intersection over Union (IOU), enhancing the model's ability to distinguish small targets and reducing scale sensitivity. This enhancement is complemented by structural modifications within the YOLOv5 network, allowing for a more nuanced focus on critical image regions.
- Application of Soft-NMS: Rather than traditional NMS, Soft-NMS better handles occlusions and overlapping targets, limiting missed and false detections in complex scenes.
- Data Augmentation Strategy: The model's generalization and adaptability to diverse environmental conditions are bolstered through data augmentation techniques like rotation, random cropping, and noise addition.

In addition to the aforementioned innovations, this study significantly contributed to the dataset by collecting 4000 real-measured sonar images from field experiments as a training dataset. This collection provides substantial data support for model training and validation and serves as a vital supplement to existing research datasets. This includes raw acoustic data from forward-looking sonar technology, sonar images, and video data, thereby facilitating further research and collaboration within the academic community.

Despite the promising results, our study has limitations. The AquaPile-YOLO algorithm has primarily been tested in controlled environments with specific water conditions, and its performance in more variable natural settings remains to be explored. Additionally, the model's computational requirements may pose challenges for real-time applications in resource-constrained environments.

The proposed AquaPile-YOLO method exhibits high applicability in marine engineering and environmental monitoring. Its ability to accurately detect underwater pile foundations can significantly enhance the efficiency and safety of harbor operations and underwater construction projects. Furthermore, the model's robustness to environmental variations makes it a promising tool for the long-term monitoring of underwater infrastructure.

Building on the foundation of the AquaPile-YOLO algorithm, future research will focus on refining and expanding capabilities for underwater pile foundation detection. The following five aspects outline the trajectory for future research and development:

- Algorithm Optimization: While the AquaPile-YOLO algorithm has demonstrated high accuracy, there is a need to continue optimizing the model structure. Reducing computational resource consumption and improving detection speed are essential to meet the demands of real-time detection, particularly in resource-constrained environments.
- Multimodal Data Fusion: To further improve detection accuracy and robustness, exploring the combination of sonar images with other sensor data, such as optical

images or LiDAR data, is a promising avenue. Multi-modal data fusion could provide a more comprehensive understanding of the underwater environment and enhance the algorithm's capabilities.

- Broader Environmental Adaptability: Assessing the model's performance across a broader range of underwater environments is crucial. Testing the algorithm in various water qualities, lighting conditions, and underwater structures will enhance the model's generality and adaptability, ensuring its effectiveness in diverse marine settings.
- Automation and Intelligence: The development of an automated sonar image collection system, integrated into underwater robots or autonomous underwater vehicles (AUV/USV/ROV/UUV), is essential for achieving fully autonomous underwater detection tasks. This advancement would increase the efficiency and safety of underwater operations.
- Engineering Application Deployment: Integrating the AquaPile-YOLO model into existing underwater monitoring systems for long-term deployment and performance evaluation is vital. Such integration will provide insights into the model's practical performance and longevity, facilitating its adoption in marine engineering and environmental monitoring projects.

Through these future directions, we expect to enhance the performance of underwater pile foundation detection technology and promote its application in the fields of marine engineering and environmental monitoring.

6. Conclusions

This paper proposes an underwater pile foundation detection method for forward-looking sonar images based on the AquaPile-YOLO algorithm. By introducing modules such as multi-scale feature fusion, attention mechanisms, and Soft-NMS, the model's detection accuracy for underwater pile foundation targets is significantly improved. The experimental results show that the AquaPile-YOLO model achieves an accuracy rate of 96.89% in underwater target identification tasks, demonstrating its efficiency and reliability in practical applications.

Author Contributions: Conceptualization, Z.X. and R.W.; methodology, Z.X., R.W. and T.C.; software, Z.X. and W.G.; validation, T.C. and Q.G.; formal analysis, R.W.; investigation, Q.G.; writing—original draft preparation, Z.X.; writing—review and editing, Z.X. and B.S.; visualization, Z.X. and W.G.; supervision, W.G.; project administration, Z.X.; funding acquisition, R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hunan Province Key Laboratory of Credible Intelligent Navigation and Positioning (funding number CINP2024002).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Acknowledgments: The authors express their sincere gratitude to China State Shipbuilding Corporation Haiying Enterprise Group Co., Ltd. for their invaluable support throughout this research endeavor. Special thanks are extended to Lei Dong, the sonar product designer, for his instrumental support in providing access to the advanced sonar equipment that was pivotal to our experimental work. We also acknowledge the contributions of Haiying Cui, whose expertise as a researcher significantly enhanced our field experiments. Additionally, we are thankful for the dedicated assistance of Hongyong Deng, whose technical prowess in equipment calibration and pool testing was crucial to the success of our research. The authors appreciate the commitment and efforts of these individuals, which played a vital role in the completion of this study.

Conflicts of Interest: Author Zhongwei Xu was employed by the company China State Shipbuilding Corporation Haiying Enterprise Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Gan, J. Development of an Underwater Detection Robot for the Structures with Pile Foundation. *J. Mar. Sci. Eng.* **2024**, *12*, 1051. [CrossRef]
2. Lu, Y.; Sang, E. Feature extraction techniques of underwater objects based on active sonars—An overview. *J. Harbin Eng. Univ.* **1997**, *18*, 43–54. (In Chinese)
3. Calder, B.R.; Linnett, L.M.; Carmichael, D.R. Bayesian approach to object detection in sidescan sonar. *IEE Proc.-Vis. Image Signal Process.* **1998**, *145*, 221–228. [CrossRef]
4. Foresti, G.L.; Gentili, S. A Vision Based System for Object Detection in Underwater Images. *Int. J. Pattern Recognit. Artif. Intell.* **2000**, *14*, 167–188. [CrossRef]
5. Guo, H. Post-Image Processing of High-Resolution Imaging Sonar. Master's Thesis, Harbin Engineering University, Harbin, China, 2002. (In Chinese).
6. Liu, C.C.; Sang, E.F. Underwater Acoustic image processing based on mathematical morphology. *J. Jilin Univ. Inf. Sci. Ed.* **2003**, *21*, 52–57. (In Chinese)
7. Kelly, J.G.; Carpenter, R.N.; Tague, J.A. Object classification and acoustic imaging with active sonar. *J. Acoust. Soc. Am.* **1992**, *91 Pt 1*, 2073–2081. [CrossRef]
8. Ye, X.F.; Zhang, Z.H.; Liu, P.X.; Guan, H.L. Sonar image segmentation based on GMRF and level-set models. *Ocean. Eng.* **2010**, *37*, 891–901. [CrossRef]
9. Wang, X. Research on Underwater Sonar Images Objective Detection and Based Respectively on MRF and Level-Set. Ph.D. Thesis., Harbin Engineering University, Harbin, China, 2010. (In Chinese).
10. Sheng, H.; Meng, F.; Li, Q.; Ma, G.; Cao, Y. Enhancement Algorithm of Side-scan Sonar Image in Curvelet Transform Domain. *Ocean. Surv. Mapp.* **2012**, *32*, 8–17. (In Chinese)
11. Sheng, Z.; Huo, G. Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning. *CAAI Trans. Intell. Syst.* **2021**, *16*, 385–392. (In Chinese)
12. Valdenegro-Toro, M. End-to-end object detection and recognition in forward-looking sonar images with convolutional neural networks. In Proceedings of the 2016 IEEE/OES Autonomous Underwater Vehicles (AUV), Tokyo, Japan, 6–9 November 2016.
13. Gong, W.; Tian, J.; Huang, H. Underwater sonar image small target recognition method based on shape features. *J. Appl. Acoust.* **2021**, *40*, 294–302. (In Chinese)
14. Bian, H.Y.; Sang, E.F.; Ji, X.C.; Zhao, J.Y. Simulation research on acoustic lens beamforming. *J. Harbin Eng. Univ.* **2004**, *25*, 43–45. (In Chinese)
15. Yang, C.Y.; Xu, F.; Wei, J.J. Seafloor sediments classification using a neighborhood gray level co-occurrence matrix. *J. Harbin Eng. Univ.* **2005**, *26*, 561–564. (In Chinese)
16. Gao, S.; Xu, J.; Zhang, P. Automatic target recognition of mine-like objects in sonar images. *Mine Warf. Ship Prot.* **2006**, *1*, 42–45. (In Chinese)
17. Fandos, R.; Zoubir, A.M.; Siantidis, K. Unified Design of a Feature-Based ADAC System for Mine Hunting Using Synthetic Aperture Sonar. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2413–2426. [CrossRef]
18. Valdenegro-Toro, M. Objectness Scoring and Detection Proposals in Forward-Looking Sonar Images with Convolutional Neural Networks. In *Artificial Neural Networks in Pattern Recognition, Proceedings of the 7th IAPR TC3 Workshop, ANNPR 2016, Ulm, Germany, 28–30 September 2016*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; pp. 209–219.
19. Zhu, K.; Tian, J.; Huang, H. Underwater objects classification method in high-resolution sonar images using deep neural network. *Acta Acust.* **2019**, *44*, 595–603. (In Chinese)
20. Sawas, J.; Petillot, Y.; Pailhas, Y. Cascade of Boosted Classifiers for Rapid Detection of Underwater Objects. In Proceedings of the 10th European Conference on Underwater Acoustics, Istanbul, Turkey, 5–9 July 2010.
21. Reed, S.; Petillot, Y.; Bell, J. Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information. *IEE Proc. Radar Sonar Navig.* **2004**, *151*, 48–56. [CrossRef]
22. Isaacs, J.C. Sonar automatic target recognition for underwater UXO remediation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015.
23. Williams, D.P.; Groen, J. A fast physics-based, environmentally adaptive underwater object detection algorithm. In Proceedings of the OCEANS 2011 IEEE—Spain, Santander, Spain, 6–9 June 2011.
24. Fandos, R.; Zoubir, A.M. Optimal Feature Set for Automatic Detection and Classification of Underwater Objects in SAS Images. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 454–468. [CrossRef]

25. Myers, V.; Fawcett, J. A Template Matching Procedure for Automatic Target Recognition in Synthetic Aperture Sonar Imagery. *IEEE Signal Process. Lett.* **2010**, *17*, 683–686. [CrossRef]
26. Hurtós, N.; Palomeras, N.; Nagappa, S.; Salvi, J. Automatic detection of underwater chain links using a forward-looking sonar. In Proceedings of the 2013 MTS/IEEE OCEANS—Bergen, Bergen, Norway, 10–14 June 2013. [CrossRef]
27. Kocak, D.M.; Dalgleish, F.R.; Caimi, F.M.; Schechner, Y.Y. A focus on recent developments and trends in underwater imaging. *Mar. Technol. Soc. J.* **2008**, *42*, 52–67. [CrossRef]
28. Fan, Z.; Xia, W.; Liu, X.; Li, H. Detection and segmentation of underwater objects from forward-looking sonar based on a modified Mask RCNN. *Signal Image Video Process.* **2021**, *15*, 1135–1143. [CrossRef]
29. Zhang, P.; Tang, J.; Zhong, H.; Ning, M.; Liu, D.; Wu, K. Self-Trained Target Detection of Radar and Sonar Images Using Automatic Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–4. [CrossRef]
30. Xie, K.; Yang, J.; Qiu, K. A Dataset with Multibeam Forward-Looking Sonar for Underwater Object Detection. *Sci. Data* **2022**, *9*, 739. [CrossRef] [PubMed]
31. Zhang, H.; Tian, M.; Shao, G.; Cheng, J.; Liu, J. Target Detection of Forward-Looking Sonar Image Based on Improved YOLOv5. *IEEE Access* **2022**, *10*, 18023–18034. [CrossRef]
32. Gaspar, A.R.; Matos, A. Feature-Based Place Recognition Using Forward-Looking Sonar. *J. Mar. Sci. Eng.* **2023**, *11*, 2198. [CrossRef]
33. Jiao, W.; Zhang, J.; Zhang, C. Open-set recognition with long-tail sonar images. *Expert Syst. Appl.* **2024**, *249 Pt A*, 123495. [CrossRef]
34. Li, Y.; Ye, X.; Zhang, W. TransYOLO: High-Performance Object Detector for Forward Looking Sonar Images. *IEEE Signal Process. Lett.* **2022**, *29*, 2098–2102.
35. Haiying Marine. HY1645 Imaging Sonar. Available online: https://www.haiyingmarine.com/index.php?a=shows&catid=74&id=106 (accessed on 15 January 2025).
36. Xia, W.; Jin, X.; Dou, F. Thinned Array Design With Minimum Number of Transducers for Multibeam Imaging Sonar. *IEEE J. Ocean. Eng.* **2017**, *42*, 892–900. [CrossRef]
37. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
38. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767v1.
39. Bochkovsky, A.; Wang, C.Y.; Liao, H.Y. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
40. Huo, G.; Wu, Z.; Li, J. Underwater Object Classification in Sidescan Sonar Images Using Deep Transfer Learning and Semisynthetic Training Data. *IEEE Access* **2020**, *8*, 47407–47418. [CrossRef]
41. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
42. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.
43. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.
44. Xing, B.; Sun, M.; Ding, M.; Han, C. Fish sonar image recognition algorithm based on improved YOLOv5. *Math. Biosci. Eng.* **2024**, *21*, 1321–1341. [CrossRef] [PubMed]
45. Qin, K.S.; Liu, D.; Wang, F.; Zhou, J.; Yang, J.; Zhang, W. Improved YOLOv7 model for underwater sonar image object detection. *J. Vis. Commun. Image Represent.* **2024**, *100*, 104124. [CrossRef]
46. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS--Improving Object Detection With One Line of Code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017. [CrossRef]

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.