Journal of *Marine Science and Engineering*

MDPI

Article

DCE-Net: An Improved Method for Sonar Small-Target Detection Based on YOLOv8

Lijun Cao [1], Zhiyuan Ma [1,*], Qiuyue Hu [1], Zhongya Xia [2] and Meng Zhao [1]

[1] The College of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China; m23181004@nue.edu.cn (L.C.); 1920191212@nue.edu.cn (Q.H.); 2420241051@nue.edu.cn (M.Z.)
[2] China Shipbuilding Industry 710th Research Institute, Yichang 443000, China
* Correspondence: zhiyuan_ma@nue.edu.cn

Abstract

Sonar is the primary tool used for detecting small targets at long distances underwater. Due to the influence of the underwater environment and imaging mechanisms, sonar images face challenges such as a small number of target pixels, insufficient data samples, and uneven category distribution. Existing target detection methods are unable to effectively extract features from sonar images, leading to high false positive rates and affecting the accuracy of target detection models. To counter these challenges, this paper presents a novel sonar small-target detection framework named DCE-Net that refines the YOLOv8 architecture. The Detail Enhancement Attention Block (DEAB) utilizes multi-scale residual structures and channel attention mechanism (AM) to achieve image defogging and small-target structure completion. The lightweight spatial variation convolution module (CoordGate) reduces false detections in complex backgrounds through dynamic position-aware convolution kernels. The improved efficient multi-scale AM (MH-EMA) performs scale-adaptive feature reweighting and combines cross-dimensional interaction strategies to enhance pixel-level feature representation. Experiments on a self-built sonar small-target detection dataset show that DCE-Net achieves an mAP@0.5 of 87.3% and an mAP@0.5:0.95 of 41.6%, representing improvements of 5.5% and 7.7%, respectively, over the baseline YOLOv8. This demonstrates that DCE-Net provides an efficient solution for underwater detection tasks.

Keywords: attention mechanism (AM); small-target detection; sonar image; YOLOv8

check for updates

Academic Editor: Weicheng Cui

Received: 5 July 2025
Revised: 26 July 2025
Accepted: 30 July 2025
Published: 31 July 2025

Citation: Cao, L.; Ma, Z.; Hu, Q.; Xia, Z.; Zhao, M. DCE-Net: An Improved Method for Sonar Small-Target Detection Based on YOLOv8. *J. Mar. Sci. Eng.* **2025**, *13*, 1478. https://doi.org/10.3390/jmse13081478

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/).

1. Introduction

As underwater detection technologies continue to evolve, sonar has become increasingly vital in various fields, including marine resource exploration [1], underwater archeology [2], military reconnaissance [3], and environmental monitoring [4]. As an imaging technology based on the principle of acoustic wave reflection, sonar can penetrate water and generate two-dimensional or three-dimensional images of underwater objects. Yet the distinctive nature of sonar imaging renders its images highly susceptible to multiple confounding factors, including noise and complex underwater environments. These issues will cause the obtained underwater sonar images to exhibit a fogging-like effect, that is, the image details become blurred, noise increases, and the perceptual quality of the images is reduced, posing significant challenges for object detection. Especially in small-target detection tasks, the small size and weak signal of the targets make them prone to being submerged by background noise, and traditional target detection methods often fail to achieve

satisfactory results. Therefore, the effective improvement of small-target detection in sonar imagery stands as a central concern within the underwater object detection domain.

Deep learning has made breakthrough progress in image processing and object detection in recent years [5]. Among them, convolutional neural networks (CNNs) have been widely used in various visual tasks due to their excellent feature learning capabilities and adaptive characteristics [6]. In object recognition, You Only Look Once (YOLO) algorithms have garnered significant attention due to their efficient real-time performance and end-to-end detection framework. YOLO algorithms innovatively transform the target recognition task into a regression problem, achieving direct mapping from image pixels to target location and category probability, thereby significantly improving detection efficiency and demonstrating clear advantages in application scenarios that require real-time processing. However, despite the remarkable success of YOLO algorithms in optical image object detection, they still face numerous challenges when processing sonar images. Additionally, small targets occupy a relatively small proportion in sonar images, and their feature information is prone to being lost, leading to higher false detection and missed detection rates in existing YOLO models when detecting small targets.

To address these challenges, researchers have attempted to improve the YOLO algorithm from multiple perspectives. On the one hand, dehazing and enhancement techniques for small-target images have garnered significant attention. Traditional image enhancement methods, such as histogram equalization, contrast stretching, and filtering techniques, can improve image quality to some extent but exhibit limited effectiveness when dealing with complex sonar images [7]. In recent years, image dehazing and enhancement algorithms based on deep learning have gradually emerged [8,9]. These methods construct CNNs to learn feature representations of images and achieve high-quality image reconstruction by optimizing network architectures and training strategies. Some studies have introduced attention mechanism (AM) to enhance features in key regions of images, thereby improving the model's perception of small targets. Moreover, feature fusion techniques have also been extensively applied in target detection tasks. By integrating multi-scale features and AM, models can more effectively extract target feature information, thereby improving detection accuracy. For example, the Squeeze-and-Excitation (SE) module optimizes feature representation through channel AM [10], while the CoordConv module enhances the expressive power of spatial features by incorporating coordinate information [11]. Although these improvements have, to some extent, enhanced the detection performance of YOLO models in complex image scenarios, conventional AM fails to effectively focus on sonar small targets with extremely low contrast and indistinct texture features.

Existing methods have not fully and synergistically addressed the compound challenges faced in sonar small-target detection, often focusing only on one or two aspects. Li et al. [12] enhanced feature extraction by introducing AM, but their approach still falls short in addressing the fogging-like issue in sonar images. Chen et al. [13] proposed an improved version based on the YOLO series object detection framework, enhancing real-time detection performance by refining multi-scale feature extraction and fusion strategies. Cao et al. [14] introduced a lightweight algorithm named MAL-YOLO, which further improved the accuracy and efficiency of underwater target detection through multi-scale feature fusion and AM. Nevertheless, most existing improvement methods are designed for optical images and fail to fully account for the unique characteristics of sonar images, such as low contrast, high noise, and dense distribution of small targets. Therefore, to better address the challenge of small-target detection in sonar images, a novel detection framework that comprehensively considers image dehazing, feature extraction, and feature fusion is needed.

To address the limitations of existing methods and systematically tackle the compound challenges of small-target detection in sonar images, this paper proposes an improved method for sonar small-target detection based on YOLOv8 named DCE-Net. DCE-Net optimizes detection performance through the synergistic action of three core modules. The Detail-Enhanced Attention Block (DEAB) [15] employs a multi-scale residual structure and channel AM to specifically remove fogging-like interference and enhance the detail textures of small targets. The lightweight Spatially Varying Convolutions module (CoordGate) [16] introduces a dynamic position-aware convolution kernel to enhance the model's sensitivity to target spatial positions, effectively reducing false alarms in complicated backgrounds. MH-EMA is an improvement of the Efficient Multi-scale Attention (EMA) [17], which performs multi-scale feature extraction and attention-weighted fusion on the input feature maps, thereby enhancing feature representation and the model's ability to capture multi-scale information. The main contributions of this paper are as follows:

(1) A distinctive end-to-end DCE-Net framework is devised to tackle the challenge of small-target detection in sonar images. Unlike previous methods, the proposed DCE-Net simultaneously enhances sonar image quality and aggregates global contextual features, thereby strengthening the feature relevance of small sonar targets.
(2) For the first time, a strategy combining sonar image defogging (DEAB) with spatial perception localization optimization (CoordGate) is proposed for underwater small-target detection. This approach efficiently extracts small-target detail feature information while suppressing background interference.
(3) A new efficient multi-scale attention module (MH-EMA) is designed. By introducing a multi-head AM, the feature fusion process is further optimized, significantly improving the model's precision and recall for small-target detection in complex backgrounds.
(4) Finally, extensive experiments validate the effectiveness and superiority of DCE-Net in the task of small-target detection in sonar images, offering a novel solution for the field of underwater target detection.

The subsequent sections are structured as follows. Section 2 reviews related work. Section 3 describes the details of the proposed model architecture. Section 4 details the datasets, implementation, and experimental results. Section 5 concludes the study and proposes future work.

2. Related Work

2.1. Image Processing

Sonar images hold significant value in fields such as marine exploration and underwater target recognition [18]. However, the sonar imaging process is easily affected by scattering and absorption effects in the water volume, resulting in hazy images that severely impact the extraction of target features and detection accuracy. Early researchers primarily relied on traditional image enhancement algorithms to improve image quality. Histogram equalization was used to expand pixel distribution and enhance contrast, but it often led to local overexposure [19]. Algorithms based on Retinex theory improved details by separating illumination components, but they were sensitive to noise [20]. Wavelet transform and guided filtering techniques suppressed noise through multi-scale decomposition, but they tended to lose edge information in complex scattering scenarios [21]. In recent years, deep learning has provided a new paradigm for sonar image dehazing [22–27]. He et al. [28] introduced a dark-channel-prior-based single-image dehazing approach, effectively addressing dehazing issues in dense fog regions, scenarios involving a low signal-to-noise ratio (SNR), and grayscale images. Cai et al. [29] introduced the Dehaze-Net architecture, which learns nonlinear mapping between hazy and clear images through an end-to-end convolutional network, effectively extracting image features and removing

haze. To address the low-SNR characteristic of hazy images, Lin et al. [30] incorporated region-aware physical constraints combined with unsupervised learning, overcoming the inherent limitations of conventional approaches that depend on paired data and struggle to adapt to real-world complex fog density distributions. Wang et al. [31] presented an unsupervised image-dehazing framework grounded in contrastive learning, leveraging the underlying feature affinities linking hazy and clear images and introducing multi-scale contrastive learning modules and adaptive feature alignment strategies to mitigate domain discrepancies caused by uneven haze distribution in real-world scenarios. Zheng et al. [32] presented Dehaze-TGGAN, a transformer-guided generative adversarial network that integrates spatial–spectral AM to tackle unpaired remote sensing image dehazing. Engin et al. [33] devised an end-to-end cycle-dehaze network that leverages cycle consistency and perceptual losses to improve texture fidelity. Sahu et al. [34] presented DCD-Net, a dual-channel deep architecture that mitigates blur and low contrast via an atmospheric scattering model. Wang et al. [35] proposed a single-image dehazing framework that restores fine details in complex scenes via multi-scale deep fusion. Ullah et al. [36] designed a novel lightweight convolutional neural network architecture named Light-DehazeNet, which resolves the issues of high computational complexity, large model parameters, and insufficient real-time performance in traditional single-image dehazing methods. Sahu et al. [37] introduced Oval-Net, a single-image dehazing method underpinned by a multi-level attention network, aiming to enhance image quality for visual measurement systems in foggy environments. Nie et al. [38] addressed the dynamic haze issues caused by complex weather and imaging conditions in real remote sensing images by proposing a dynamic dehazing method based on contrastive haze-aware learning. However, most existing methods are designed for optical images and inadequately account for the traits of sonar images, including fan-shaped distortion and speckle-noise distributions, leading to significant artifact retention after dehazing.

2.2. YOLO Series Algorithms

The YOLO architecture has established itself as the benchmark in the field of target detection due to its single-stage detection framework and real-time performance advantages. YOLOv1 pioneered the transformation of the detection task into a grid cell regression problem but suffered from missed detections of small targets [39]. YOLOv3 introduced multi-scale prediction and the Darknet-53 backbone network, significantly improving adaptability to targets of varying sizes [40]. YOLOv4 achieved a balance between accuracy and speed through Mosaic data augmentation and Cross Stage Partial Darknet (CSPDarknet) structural optimization [41], while the flexible architecture design of YOLOv5 led to its rapid adoption in industrial applications. YOLOv8 employs dynamic label assignment strategies and hybrid AM, achieving a 51.2% Average Precision (AP) on the MS COCO dataset. However, in the field of underwater object detection, the YOLO series faces significant challenges. Cao et al. [42] improved tracking accuracy and real-time performance in complex underwater scenes using an enhanced YOLOv3 algorithm. Lu et al. [43] devised a streamlined YOLO architecture for real-time single-class object detection. Li et al. [44] advanced an enhanced YOLOv3 framework tailored to ship detection, addressing the challenge of ship recognition in complex backgrounds within remote sensing images. Yang et al. [45] introduced a lightweight deep learning model, SS-YOLO, which improved feature fusion, incorporated Multi-Head Self-Attention mechanisms (MHSA), and utilized residual connections to tackle the issue of low detection accuracy in complex underwater scenes. Ge et al. [46] developed an improved YOLOv5-based method for orientation-sensitive object detection, which effectively boosts the performance for detecting multi-scale, orientation-sensitive targets in complex scenes through sample grouping

strategies and network structure optimization. Cao et al. [14] developed a lightweight detection algorithm, MAL-YOLO, based on an improved YOLOv5 model, to tackle challenges such as large-target scale variations, complex background interference, and limited computational resources in side-scan sonar image target detection. Liu et al. [47] tackled the degradation of detection performance caused by noise, blur, and rotational jitter in real-world scenarios by proposing a degradation model construction strategy combining YOLO networks with traditional image processing methods. Wang et al. [48] proposed an improved YOLO model integrating multi-scale feature fusion and AM to enhance the extraction of shape and texture features of blurred underwater targets. Although these improved methods have enhanced the performance of underwater target detection to some extent, they overlook the issues of the small number of pixels and abundant fogging-like interference in underwater small targets, and the image quality of small targets still cannot be improved.

2.3. Feature Fusion Technology

Feature fusion techniques, by integrating feature information from different levels or modalities, have become crucial for enhancing detection performance. In hierarchical fusion, the Feature Pyramid Network (FPN) constructs a multi-scale feature pyramid through top-down lateral connections, effectively addressing the issue of target scale variation [49]. Furthermore, a Path Aggregation Network (PANet) enhances localization information flow by adding a bottom-up branch [50], while a Bidirectional FPN (BiFPN) achieves efficient multi-scale fusion through learnable weights [51]. To address the challenges posed by seabed reverberation noise, a low foreground pixel ratio, and poor imaging resolution in sonar image target detection, Wang et al. [52] introduced MLFFNet, a multi-level feature fusion network that enhances the localization and classification of targets in sonar images by fusing multi-scale features and channel AM. Li et al. [53] introduced a dynamic feature aggregation framework, DG-FPN, based on Graph Convolutional Networks (GCNs), leveraging the dynamic modeling capabilities of graph convolutions to boost the precision of detecting multi-scale targets. Tong et al. [54] proposed a novel FPN based on feature fusion to improve the detection performance of small targets in infrared images. Zhang et al. [55] introduced a drone-based hyperspectral image classification method that fuses shallow and deep features, combining spectral features with abstract features extracted by deep convolutional networks to improve classification performance. Chen et al. [56] addressed the challenge of poor detection performance for small targets in remote sensing imagery by proposing an improved bidirectional cross-scale connected feature fusion network. Zhang et al. [57] developed a hierarchical feature integration network with an AM, termed MFANet, which captures multi-scale features using the Deep Atrous Spatial Pyramid Pooling (DASPP) module and enhances feature representation and spatial-channel contextual dependencies through the Feature Alignment Fusion (FAF) module and Contextual Attention (CA) module. Wang et al. [58] introduced a single-stage detection framework leveraging CNNs, integrating techniques for extracting and fusing features across scales to detect and recognize targets in Synthetic Aperture Radar (SAR) images. Hu et al. [59] developed a Multi-Level Adaptive Attention Fusion Network (MLAAF) to address issues of information redundancy or loss of critical information caused by modal feature differences and multi-scale feature inconsistencies in infrared and visible image fusion. Liang et al. [60] tackled the insufficient fusion of spectral and spatial features in hyperspectral image classification tasks by proposing a Deep Multi-scale Spectral-Spatial Feature Fusion Network (DMFNet). Yu et al. [61] introduced GLF-Net, a target detection method for remote sensing aircraft images, which addresses the challenges of aircraft target detection in complex backgrounds through the fusion of global and local features at multiple scales.

However, the aforementioned methods focus merely on one aspect of small-target detection, failing to comprehensively consider defogging image processing and global information fusion, which leads to prominent issues of missed and false detections.

3. Method

3.1. Overall Network Architecture

The specific architecture is illustrated in Figure 1. The enhanced YOLOv8 model comprises three main components. The Backbone is tasked with extracting features from the input image, and the Neck performs multi-scale feature fusion to enrich the representation, while the Head is responsible for performing target detection based on the extracted features. Within the Backbone, the DEAB and CoordGate modules are introduced to improve the robustness of feature extraction. Within the Head, the MH-EMA module is added to enhance the global feature correlation. In the actual network architecture, the DEAB and CoordGate modules are sequentially integrated before the SPPF module in the Backbone. The DEAB not only improves feature extraction through its depth-enhanced AM but also demonstrates remarkable performance in processing blurred images, particularly in defogging tasks. In sonar image processing, where the distinction between target and background information is often subtle, and issues such as weak target pixels and small target sizes are prominent, the introduction of the DEAB significantly addresses these challenges. By leveraging its depth-enhanced AM, the model can more effectively highlight fine-grained features within the image, thereby improving the detection capability for small-sized targets. Furthermore, the application of the DEAB in defogging indicates its ability to improve image quality and enhance target visibility, which is crucial for target recognition and localization in sonar images. Following the DEAB, we further introduce the CoordGate module to enhance the model's ability to capture spatially varying features. The CoordGate module is a lightweight structure that incorporates a coordinate encoding network and a multiplicative gating mechanism, enabling the convolutional neural network to selectively amplify or attenuate filters based on spatial location. By utilizing geometric cues, the CoordGate module can more accurately localize and identify targets in the image. Even in cases of partial occlusion, it provides additional cues to infer the complete shape of the target, thereby improving detection accuracy and reducing false positive rates. To capture pixel-level information of the target images, we add the MH-EMA module in the first layer of the Neck after the SPPF module. The key of the MH-EMA lies in the introduction of the multi-head attention, which is proficient at capturing long-range dependencies between different positions in the feature map. Meanwhile, by adaptively modulating the weights of features across scales, the process of feature pyramid fusion is further optimized.

3.2. Core Modules

3.2.1. DEAB Module

The DEAB module integrates convolutional operations with AM to achieve pixel expansion and structural repair in images, as shown in Figure 2. Its core consists of Detail-Enhanced Convolution (DEConv) and Content-Guided Attention (CGA) mechanism. Detail-Enhanced Convolution strengthens feature extraction capabilities by using ordinary convolution and differential convolution in parallel. Differential convolution includes Center Difference Convolution (CDC), Angular Difference Convolution (ADC), Horizontal Difference Convolution (HDC), and Vertical Difference Convolution (VDC). These opera-

tions embed traditional local descriptors within the convolutional layer, thereby enhancing the feature depiction and overall robustness. The formula is provided below.

$$VC(x) = \sum_{i=1}^{k}\sum_{j=1}^{k} X_{i,j}\cdot W_{i,j}. \tag{1}$$

$$CDC(x) = \sum_{i=1}^{k}\sum_{j=1}^{k} \left(X_{i,j} - X_{c,c}\right)\cdot W_{i,j}. \tag{2}$$

$$ADC(x) = \sum_{\theta=0^\circ}^{180^\circ} Conv(X, W_\theta). \tag{3}$$

$$HDC(x) = \sum_{i=1}^{k}\sum_{j=1}^{k} W_{i,j}\cdot \sum_{i=1}^{h}\sum_{j=1}^{w} (X_{i,1} - X_{i,w}). \tag{4}$$

$$VDC(x) = \sum_{i=1}^{k}\sum_{j=1}^{k} W_{i,j}\cdot \sum_{i=1}^{h}\sum_{j=1}^{w} (X_{1,j} - X_{h,j}). \tag{5}$$

$$DEconv(X) = VC(x) + CDC(x) + ADC(x) + HDC(x) + VDC(x) \tag{6}$$

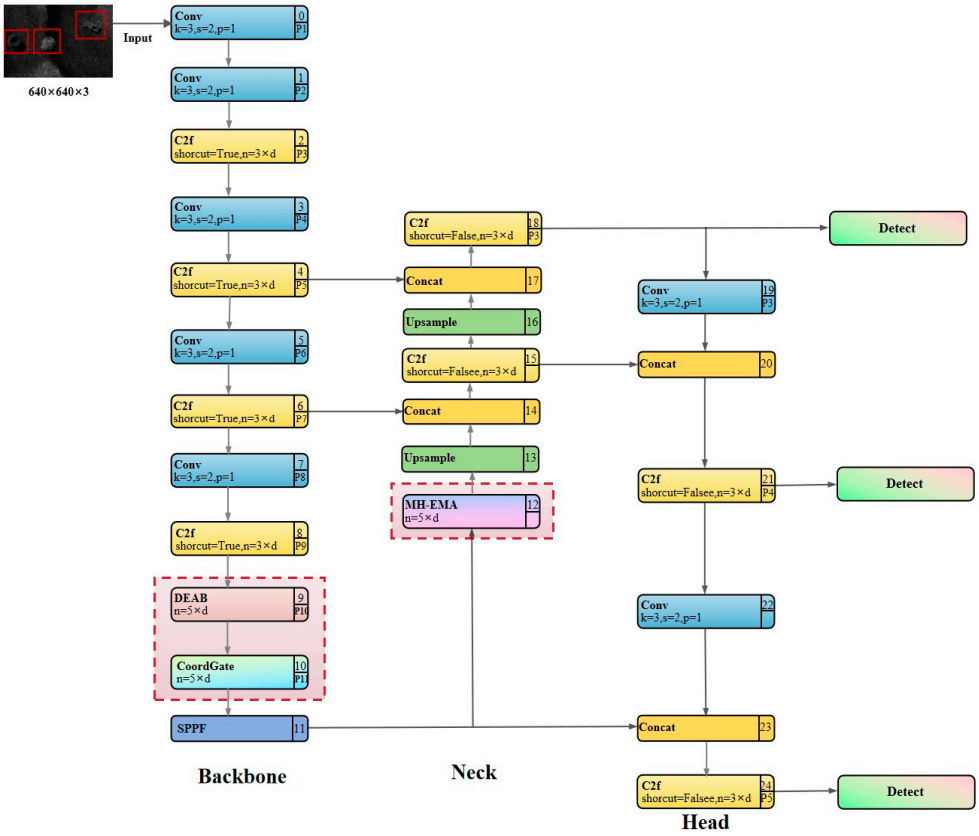


Figure 1. Overall framework of DCE-Net, including DEAB, CoordGate and MH-EMA module. Red dashed boxes indicate newly added modules.

Here, k signifies the size of the convolution kernel, with $X_{i,j}$ denoting the pixel value of the input feature map at coordinates (i, j); $W_{i,j}$ indicates the positional weight of the convolution kernel; $X_{c,c}$ represents the pixel value at the center position of the input feature map; h and w, respectively, represent the total numbers of rows and columns of the input feature map (both are integers), and W_θ denotes the weight in the direction θ of the convolution kernel.

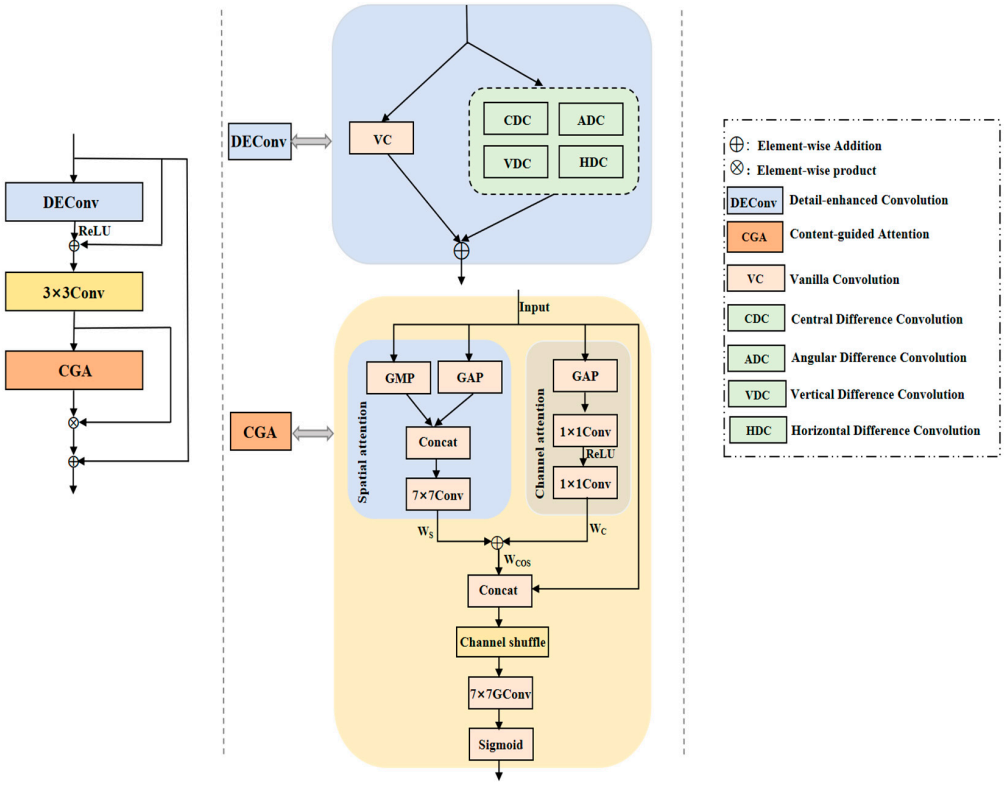


Figure 2. Structure diagram of the DEAB module.

The CGA module functions as a two-stage attention generator, creating spatial attention maps via a coarse-to-fine refinement process. Specifically, it initially generates a coarse spatial attention map, which is subsequently refined by each channel of the input feature map, ultimately obtaining a channel-specific Spatial Importance Map (SIM). Let $X \in \mathrm{R}^{C\times H\times W}$ represent the input features from the upper layer, and the formula is as follows:

$$W_c = C_{1\times1}\left(max\left(0, C_{1\times1}\left(X^C_{GAP}\right)\right)\right). \tag{7}$$

$$W_S = C_{7\times7}\left(\left[X^S_{GAP}, X^S_{GMP}\right]\right). \tag{8}$$

where max $(0, x)$ denotes the ReLU activation function, and $C_{1\times1}(\cdot)$ represents a convolution with a kernel size of 1×1. X^C_{GAP} indicates the feature obtained by performing Global Average Pooling (GAP) along the channel dimension, while X^S_{GAP} and X^S_{GMP}. represent the features obtained by performing GAP and Global Max Pooling (GMP) along the spatial dimension, respectively. The coarse SIM is obtained through matrix addition operations, and the result is

$$W_{cos} = W_c + W_s. \tag{9}$$

Finally, after the operations of concatenation, channel shuffling, and group convolution, the refined SIM is obtained, and the result is

$$W = \sigma(GC_{7\times7}(CS([X, W_{cos}]))). \tag{10}$$

where σ refers to the Sigmoid activation function, $CS(\cdot)$ represents channel shuffling, and $GC_{7\times7}(\cdot)$ indicates a group convolution layer with a kernel size of 7×7.

3.2.2. CoordGate Module

The CoordGate module, by integrating a multiplicative gate with a coordinate encoding network, selectively modulates the output of the convolutional layer to perform spatially variant convolution operations, as shown in Figure 3. First, the input data is processed through a standard convolutional block to generate global convolutional channels. Subsequently, these channels are multiplied element-wise by a gate mask of the same size, similar to an AM, to produce locally variant convolutional outputs. The formula can be expressed as

$$y = h(x) \odot (g(c) \cdot g(h(x))). \tag{11}$$

where y is the output, and $\odot$ refers to the Hadamard product. $h(x)$ represents the processed result of the input data x through a standard convolutional block, $g(c)$ is the processed result of the coordinates c through a fully connected encoding network, and $g(h(x))$ is the processed result of $h(x)$ through a fully connected encoding network.

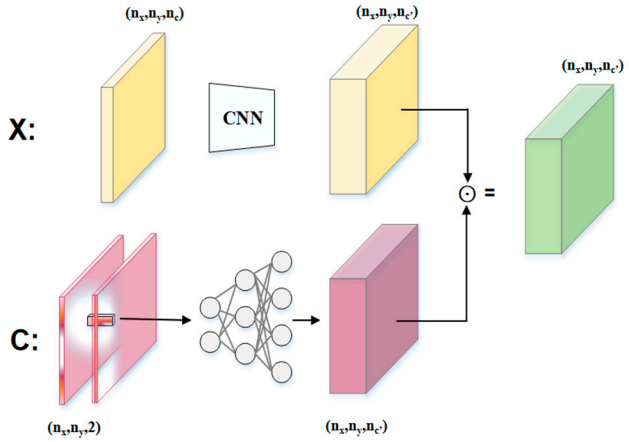


Figure 3. Structure diagram of the CoordGate module.

3.2.3. MH-EMA Module

The MH-EMA module not only achieves feature pyramid fusion through spatial learning but also introduces a multi-head AM and dynamic weight adjustment to further optimize the fusion process. This makes the module more efficient and flexible when dealing with complex image data. The specific structure of MH-EMA is shown in Figure 4. Suppose the input feature map is $X \in R^{C \times H \times W}$, where C is the number of channels, H is the height, and W is the width of the feature map. With each grouped vector $X_i \in R^{C/G \times H \times W}$, the convolutional layer maps the grouped feature map to Q, K, and V, so we have

$$Attention = softmax\left(\frac{QK^T}{\sqrt{\frac{C}{(G \times num_heads)}}}\right) \cdot V. \tag{12}$$

Here, the query vector is $Q \in R^{t_{q \times d_k}}$, the key vector is $K \in R^{t_{k \times d_k}}$, and the value vector is $W \in R^{t_{k \times d_v}}$. d_k and d_v are the dimensions of the keys and values for each head, respectively, and are typically set to $d_k = d_v = d_{model}/h$ to maintain the total dimensionality across multiple heads. d_{model} is the dimensionality of the input and output, usually fixed as the model's hidden dimension, and h is the number of heads.

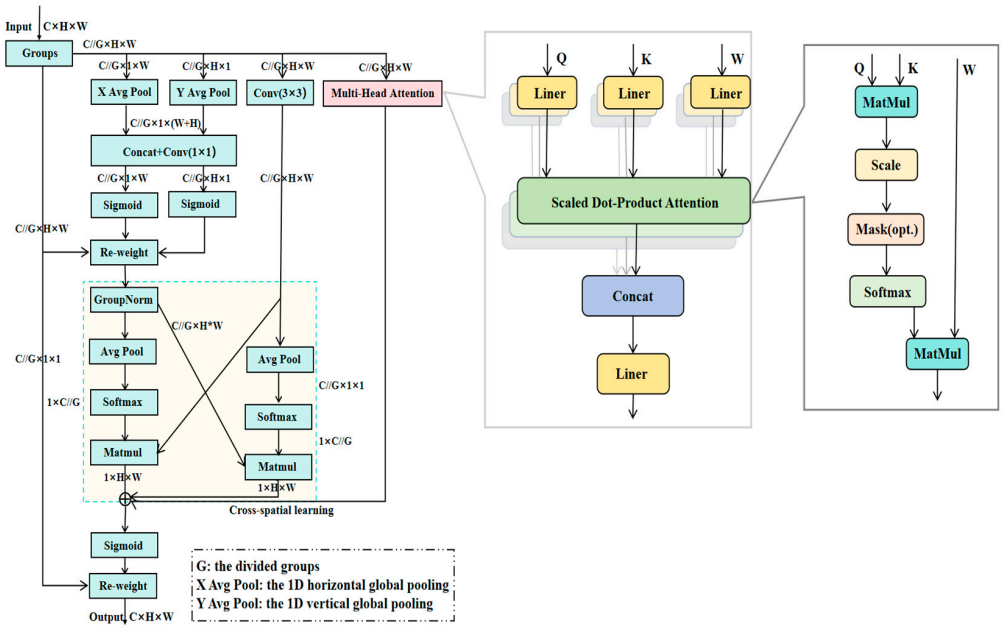


Figure 4. Structure diagram of the MH-EMA module. The green section on the left represents the original EMA module.

The calculation formula for each head is as follows:

$$MultiHead(Q, K, V) = Concat(head_1, head_2 \cdots head_h)w^O. \tag{13}$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right). \tag{14}$$

Here, $W_i^Q \in R^{d_{model} \times d_k}$, $W_i^K \in R^{d_{model} \times d_k}$, and $W_i^V \in R^{d_{model} \times d_v}$ are learnable parameters, and $W^O \in R^{h \cdot d_v \times d_{model}}$ is the output weight.

$$X_{GAP} = \frac{1}{W}\sum_{w=1}^{W} X_i. \tag{15}$$

$$Y_{GAP} = \frac{1}{H}\sum_{h=1}^{H} X_i. \tag{16}$$

$$Re - weight(X_i) = \sigma(Conv(Concat(X_{GAP}, Y_{GAP}))) \cdot X_i. \tag{17}$$

The final output feature map is as follows:

$$y = \sum_{i=1}^{G} Re - weight(X_i). \tag{18}$$

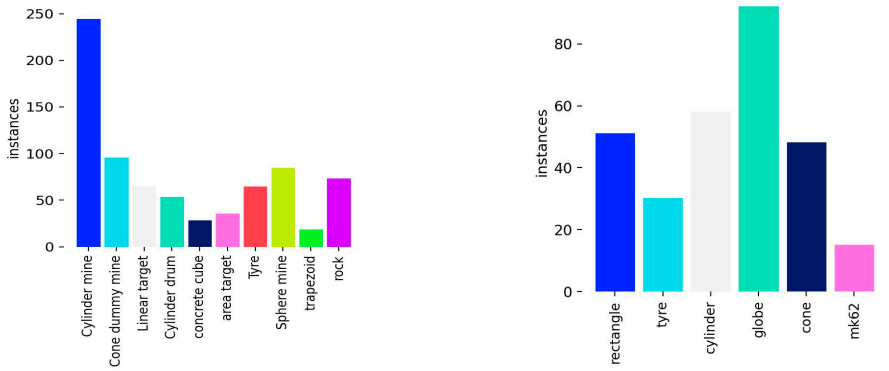
4. Experiments and Results

In this section, we first introduce the dataset, experimental setup, and evaluation metrics utilized in the experiments. Subsequently, ablation studies and visual analysis are conducted. Finally, we compare our method with other advanced target detection methods for sonar images.

4.1. Design of Experiments

Due to the limitation of scarce publicly available sonar datasets for single-image multi-target detection, this study chose to conduct training and validation of relevant experiments on the self-constructed datasets in the laboratory, the SD-Dataset and SQ-Dataset, respectively. The SD-Dataset is constructed by synthesizing real seafloor backgrounds

and real target samples. The background images are obtained from the seafloor by an interferometric synthetic aperture sonar prototype developed in our laboratory during sea trials. The typical targets are segmented from real sonar images. After synthesis using simulation software, the dataset is further processed with data augmentation and noise addition. SQ-Dataset is a real dataset collected by an interferometric synthetic aperture sonar at a frequency of 240 kHz. The SD-Dataset comprises target images of 10 classes under underwater scenarios, including cylinder mine, cone dummy mine, linear target, cylinder drum, concrete cube, area target, tire, sphere mine, trapezoid, and rock. The SQ-Dataset contains images of six classes of underwater targets resembling mines, including tire, cylinder, globe, cone, mk62, and rectangle. Both datasets cover targets of various sizes and shapes. Figure 5 illustrates the distribution of the number of targets in each class of the datasets. Some samples of the datasets are shown in Figure 6. The SD-Dataset is partitioned into a training set and a validation set, containing 319 and 103 images, respectively. The SQ-Dataset is partitioned into a training set, a test set, and a validation set comprising 209, 103, and 91 images, respectively.



(**a**) SD-Dataset (**b**) SQ-Dataset

Figure 5. The distribution of the number of various types of targets.

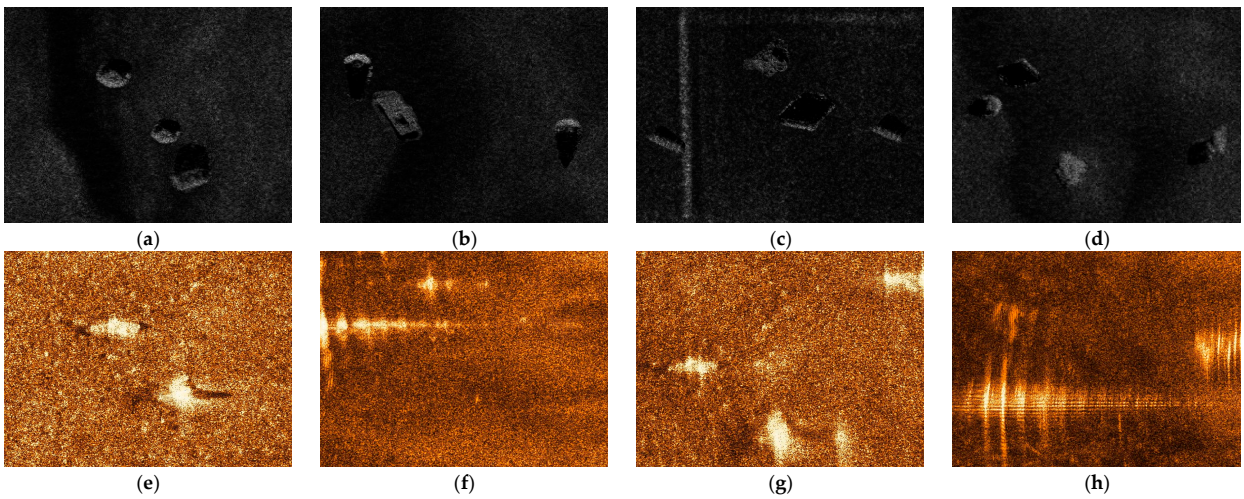


Figure 6. Some sample images of the SD-Dataset and SQ-Dataset. (**a**–**d**) represent a portion of the SD-Dataset, while (**e**–**h**) represent a portion of the SQ-Dataset.

In the experiment, DCE-Net was trained and tested using the PyTorch 2.0.1 framework on a system equipped with an AMD Ryzen 9 7495HX CPU @ 2.5 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX 4060 GPU, running a 64-bit Windows 10 operating system with CUDA 11.3. All input images were cropped and resized to 640 × 640 pixels.

For evaluation purposes, we employed recall, precision, F1_score, AP, and mean average precision (mAP) as the performance metrics for the model. The calculation formulas are as follows:

$$Recall = \frac{T_P}{(T_P + F_N)}. \tag{19}$$

$$Precision = \frac{T_P}{(T_P + F_P)}. \tag{20}$$

$$F1_score = 2 \times \frac{Recall \times Precision}{Recall + Precision}. \tag{21}$$

$$AP = \int_0^1 P(r)dr. \tag{22}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{n} AP_i. \tag{23}$$

Here, *mAP@0.5* denotes the mAP, with the Intersection over Union (IoU) threshold set to 0.5. *mAP@0.5:0.95* refers to the mAP calculated with IoU thresholds ranging from 0.5 to 0.95. T_P stands for True Positive, F_P for False Positive, and F_N for False Negative. $P(r)$ refers to the function graph bounded by recall and precision. N refers to the number of object classes. By integrating the recall and precision values obtained from each detection, a precision–recall (P-R) curve is constructed, and the AP is computed using interpolation methods.

4.2. Ablation Studies

Among various versions of YOLO models, YOLOv8m demonstrates a favorable balance between accuracy and speed with a comparable number of parameters [62]. In this article, YOLOv8 is employed as the baseline model for further improvements. Subsequently, the DEAB and the CoordGate are integrated into the Backbone of YOLOv8, and the MH-EMA is incorporated into the Neck network to construct DCE-Net.

To assess the contributions of each module within DCE-Net, ablation experiments are conducted on the SD-Dataset. Throughout the training of all models, we employed the auto-optimizer provided in the YOLOv8 framework for adaptive parameter configuration. The hyperparameter settings were strictly unified across all models. Specifically, the initial learning rate was set to 0.01 and kept constant during the entire training process, resulting in a final learning rate of 0.01. The momentum coefficient was fixed at 0.937, and weight decay was set to 5×10^{-4}. To stabilize the parameter updates in the early training stages, a three-epoch warmup strategy was adopted in which the momentum was initialized to 0.8 and the bias learning rate was set to 0.1. The total number of training epochs was 260, with a batch size of 3. As is shown in Table 1, the experimental results of all models are based on the best result obtained from the two independent runs (with the same parameter settings, only the random seeds were different).

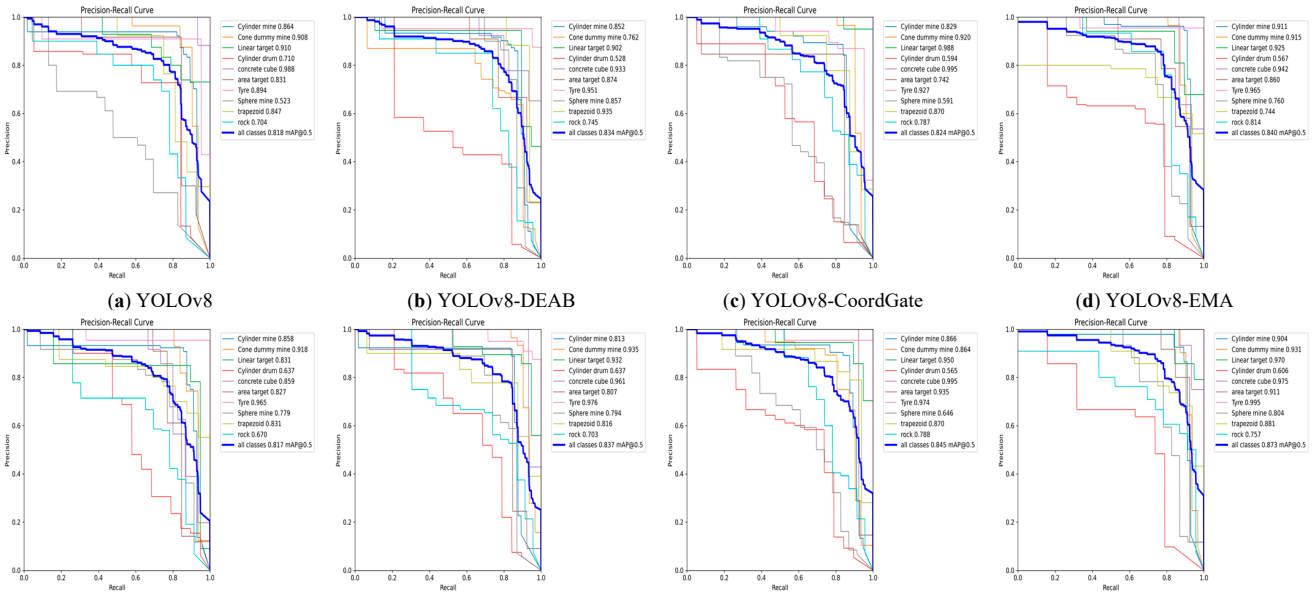
Table 1 presents the quantitative results of the ablation study conducted on the YOLOv8 model using the SD dataset. The results in the table indicate that the performance of the baseline model significantly improved across multiple evaluation metrics with the introduction of various components. Notably, the introduction of the EMA module enhanced the mAP@0.5:0.95 and mAP@0.5 metrics from 33.9% and 81.8% to 36.7% and 84%, respectively, demonstrating the effectiveness of the EMA module in boosting model performance. Additionally, the incorporation of the DEAB and CoordGate modules also led to improvements in the mAP@0.5:0.95 metric by 2.7% and 1.7%, respectively, while optimizing the model's recall and F1_score without significantly increasing the number of parameters or computational load.

Table 1. Quantitative evaluation of ablation study with different components on the SD-Dataset.

Model	mAP@0.5:0.95 (%)	mAP@0.5 (%)	R (%)	F1 (%)	Params (M)	FLOPs (G)	FPS (f·s^{-1})
YOLOv8	33.9	81.8	81.2	80	3	8.1	277
YOLOv8-DEAB	36.6 (+2.7)	83.4 (+1.6)	76.8	74	3.35	8.3	263
YOLOv8-CoordGate	35.6 (+1.7)	82.4 (+0.6)	81.7	81	3.01	8.1	250
YOLOv8-EMA	36.7 (+2.8)	84 (+2.2)	84	82	3.01	8.1	277
YOLOv8-DEAB-CoordGate	38.7 (+4.8)	83.8 (+2)	78.2	78	3.35	8.3	200
YOLOv8-DEAB-EMA	38.5 (+4.6)	83.3 (+1.5)	76.9	76	3.35	8.4	217
YOLOv8-CoordGate-EMA	37.7 (+3.8)	84.5 (+2.7)	81.8	80	3	8.1	263
YOLOv8-DEAB-CoordGate-EMA	39.7 (+5.8)	86.8 (+5)	82	80	3.35	8.4	250
Ours (DCE-Net)	41.6 (+7.7)	87.3 (+5.5)	83.5	81	5.19	9.7	217

Further experimental results suggest that combining the DEAB, CoordGate, and EMA modules can further enhance model performance. For instance, the YOLOv8-DEAB-CoordGate-EMA model achieved mAP@0.5:0.95 and mAP@0.5 scores of 39.7% and 86.8%, respectively, surpassing the performance of combinations using only two of these modules. This indicates a synergistic optimization effect among these modules, collectively improving the model's detection capabilities.

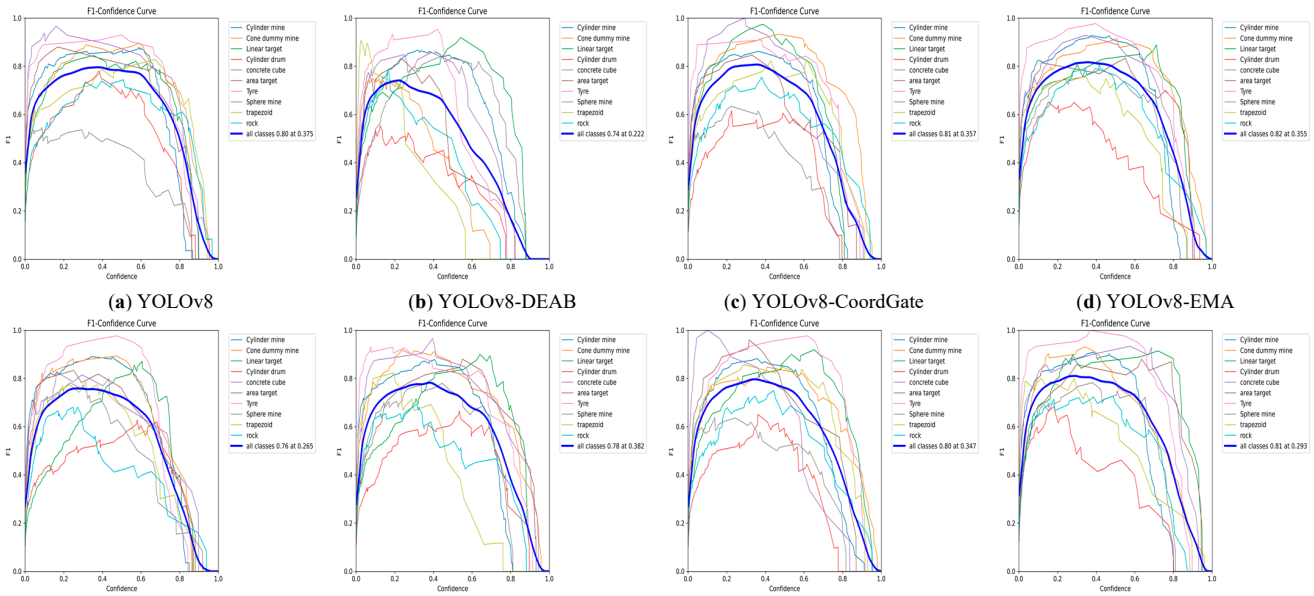
It is worth noting that our proposed DCE-Net model demonstrated outstanding performance across all evaluation metrics, with mAP@0.5:0.95 and mAP@0.5 reaching 41.6% and 87.3%, respectively, significantly outperforming other models. Although DCE-Net has a relatively higher number of parameters and Floating Point Operations per Second (FLOPs), its advantages in precision and recall indicate that the model is more efficient and accurate in dealing with multifaceted tasks. Crucially, DCE-Net achieves 41.6% mAP@0.5:0.95 while sustaining an inference speed of 217 f·s^{-1}. This demonstrates that the synergistic design of DEAB, CoordGate, and MH-EMA attains an excellent balance between accuracy enhancement and frame rate preservation, ensuring real-time deployment capability even in complex scenarios. The visualization results of various models in terms of P–R curves and F1_score curves can be observed in Figures 7 and 8.



(**a**) YOLOv8 (**b**) YOLOv8-DEAB (**c**) YOLOv8-CoordGate (**d**) YOLOv8-EMA

(**e**) YOLOv8-DEAB-CoordGate (**f**) YOLOv8-DEAB-EMA (**g**) YOLOv8-CoordGate-EMA (**h**) Ours (DCE-Net)

Figure 7. P–R curves of different methods on the SD-Dataset.



(**a**) YOLOv8 (**b**) YOLOv8-DEAB (**c**) YOLOv8-CoordGate (**d**) YOLOv8-EMA

(**e**) YOLOv8-DEAB-CoordGate (**f**) YOLOv8-DEAB-EMA (**g**) YOLOv8-CoordGate-EMA (**h**) Ours (DCE-Net)

Figure 8. F1_score curves of different methods on the SD-Dataset.

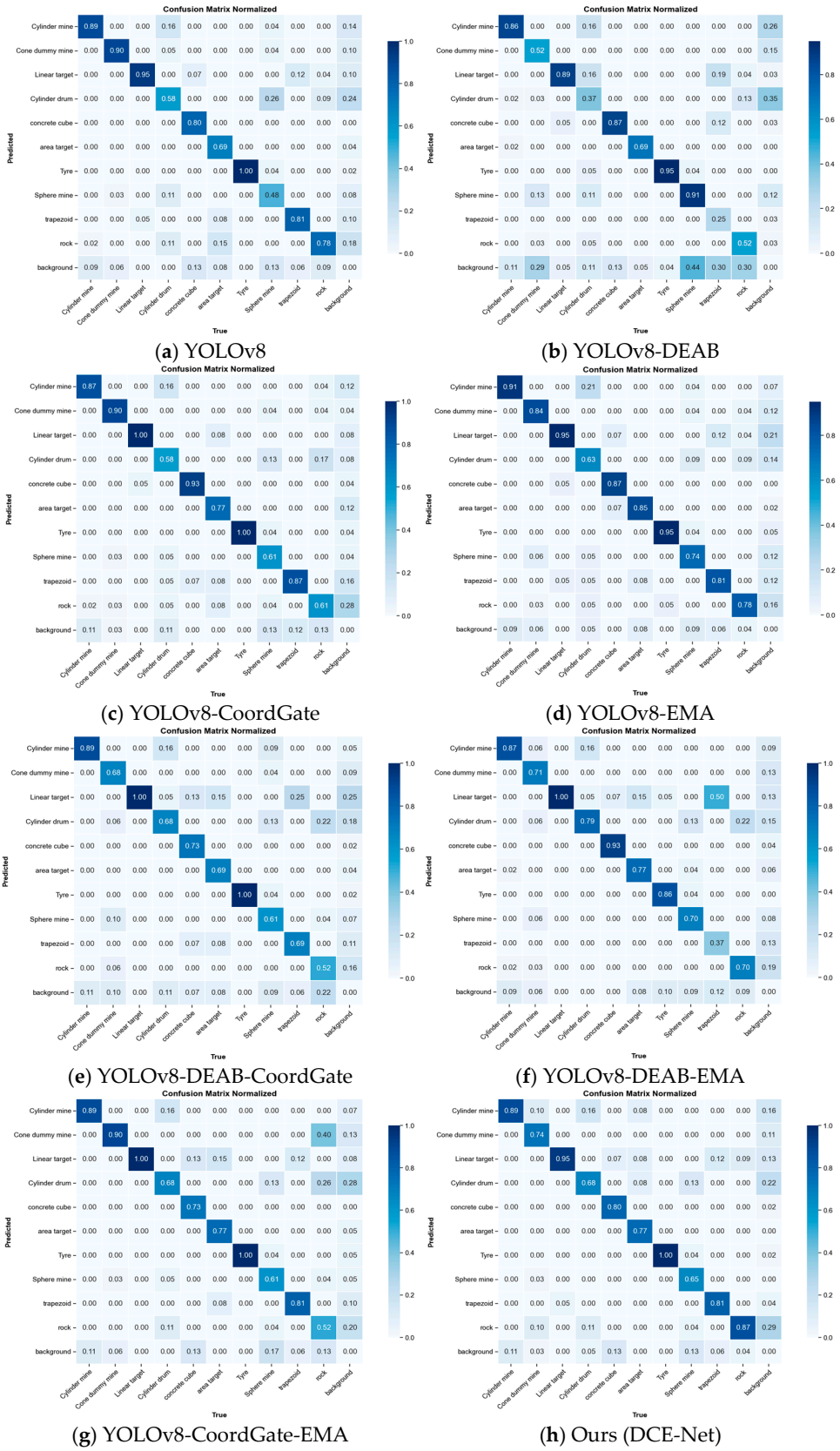
To further demonstrate the generalization performance of DCE-Net, we also conducted the same experiments on the SQ-Dataset. The results are shown in Table 2. From Table 2, the base model YOLOv8 achieved only 87.3% and 43.9% on the mAP@0.5 and mAP@0.5:0.95 metrics, respectively. Our proposed DCE-Net model performed excellently across all evaluation metrics, with the mAP@0.5 and mAP@0.5:0.95 reaching 92.2% and 49.5%, respectively, significantly outperforming other models. Through quantitative evaluation, the superior performance of DCE-Net in object detection tasks was further confirmed.

Table 2. Quantitative evaluation of ablation study with different components on the SQ-Dataset.

Model	mAP@0.5:0.95 (%)	mAP@0.5 (%)	R (%)	F1 (%)	Params (M)	FLOPs (G)	FPS (f·s^{-1})
YOLOv8	43.9	87.3	77.1	80	3	8.1	143
YOLOv8-DEAB	45.6 (+1.7)	87.1 (−0.2)	80.7	83	3.35	8.3	113
YOLOv8-CoordGate	46.2 (+2.3)	89.2 (+1.9)	81.8	84	3.01	8.1	121
YOLOv8-EMA	46.3 (+2.4)	87.1 (−0.2)	83.5	82	3.01	8.1	111
YOLOv8-DEAB-CoordGate	47.7 (+3.8)	87.8 (+0.5)	79.6	82	3.35	8.3	108
YOLOv8-DEAB-EMA	47.6 (+3.7)	89.9 (+2.6)	82.3	85	3.35	8.4	135
YOLOv8-CoordGate-EMA	46.6 (+2.7)	90.4 (+3.1)	82	84	3	8.1	142
YOLOv8-DEAB-CoordGate-EMA	48.4 (+4.5)	89.9 (+2.6)	83.6	84	3.35	8.4	147
Ours (DCE-Net)	49.5 (+5.6)	92.2 (+4.9)	83.1	88	5.19	9.7	125

4.3. Visual Analytics

After being trained and tested on the SD-Dataset, DCE-Net achieves an mAP@0.5 of 86.8%, demonstrating high-precision detection of small targets underwater. Figure 9 illustrates the confusion matrices for various models on the SD-dataset, which are used to evaluate the classification performance of each model in target detection tasks. In the confusion matrices, larger values along the main diagonal indicate higher detection accuracy. Non-zero values in regions outside the main diagonal represent false detections, with larger values corresponding to higher false detection rates.



(**a**) YOLOv8 (**b**) YOLOv8-DEAB

(**c**) YOLOv8-CoordGate (**d**) YOLOv8-EMA

(**e**) YOLOv8-DEAB-CoordGate (**f**) YOLOv8-DEAB-EMA

(**g**) YOLOv8-CoordGate-EMA (**h**) Ours (DCE-Net)

Figure 9. Confusion matrix of different methods on the SD-Dataset.

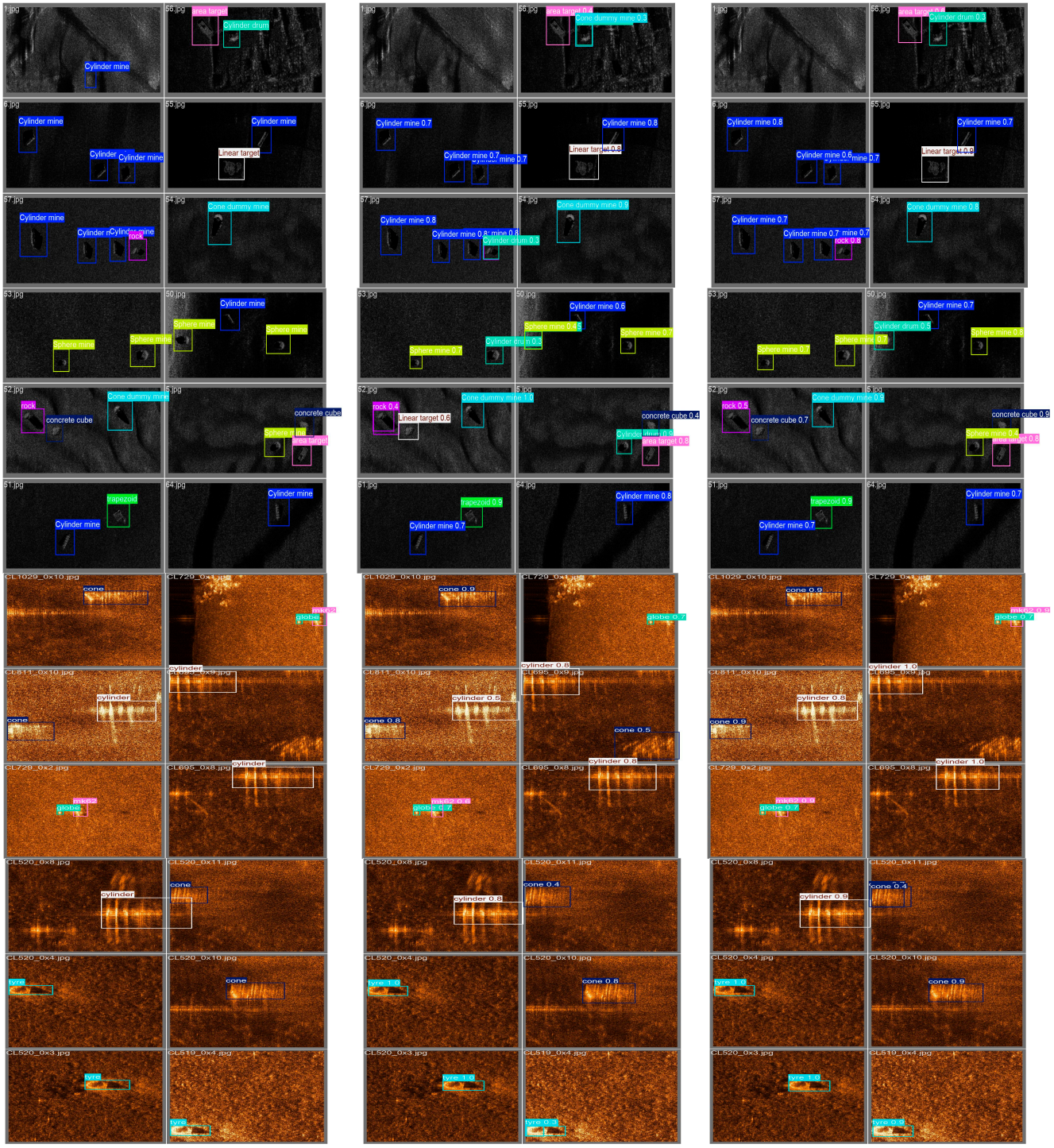
It can be seen from the confusion matrix that the model in Figure 9a exhibits certain limitations in handling small targets in complex underwater scenarios, particularly showing low recognition accuracy for small targets such as sphere mines, with a rate of only

48%. The introduction of the DEAB in Figure 9b improves detection accuracy for certain categories, but the false detection rate remains relatively high, indicating that the DEAB may introduce some noise while enhancing detail information. The module in Figure 9c effectively reduces the false detection rate by incorporating the CoordGate module, but the recall rate for some categories decreases, suggesting that this module may sacrifice some detection sensitivity while reducing false detections. The model in Figure 9d significantly improves detection accuracy and recall rates by introducing the EMA module, particularly excelling in target detection tasks within complex backgrounds, demonstrating that the cross-dimensional AM effectively captures pixel-level feature information. The model in Figure 9e further enhances detection accuracy by combining the DEAB and CoordGate modules, particularly performing well in small-target detection, indicating that the synergistic effect of the two modules effectively enhances detail information and reduces false detections. The model in Figure 9f performs excellently across multiple categories, particularly in target detection tasks within complex backgrounds, demonstrating that the combination of detail enhancement and cross-dimensional AM significantly improves model performance. The model in Figure 9g maintains high detection accuracy and recall rates while reducing false detections by combining the CoordGate and EMA modules, indicating that the synergistic effect of spatially variant convolution and AM effectively enhances model robustness. The model in Figure 9h performs outstandingly across all categories, demonstrating that the synergistic effect of the DEAB, CoordGate, and MH-EMA modules significantly improves detection accuracy and robustness. In summary, DCE-Net, through multi-module collaborative optimization, demonstrates superior performance in underwater target detection tasks within complex environments, particularly excelling in small-target detection and reducing false detections.

Figure 10 presents the visualization of detection results of the base model YOLOv8 and DCE-Net on the SD-Dataset and SQ-Dataset. Through comparative analysis, the performance differences between the two models in underwater object detection tasks can be intuitively evaluated. It is evident from this figure that the base model YOLOv8 exhibits significant false detection in the first two sets of images, false alarms in the third set, and relatively low detection accuracy in the fourth set. These results indicate that the YOLOv8 model still has limitations in detecting small-target and sophisticated underwater backgrounds, making it prone to false detections and false alarms. In contrast, DCE-Net, by considering image dehazing and feature extraction and integrating three types of modules, significantly enhances the model's predictive performance, thereby improving detection accuracy and markedly reducing the false detection rate.

4.4. Comparative Experiments

To demonstrate the superior performance of the proposed DCE-Net framework in sonar small-target detection, this section presents the results of comparative experiments with several classical object detection models, including Faster R-CNN [63], RetinaNet [64], YOLOv8m, YOLOv9m, YOLOv10m, and YOLO11s. All models were trained and tested using the self-constructed sonar image dataset, the SD-Dataset. All the basic parameter settings for the models, such as batch size, learning rate, and number of training epochs, remain consistent with those described previously. Training was halted when the model reached convergence. The experimental results are presented in Table 3.



Labeled images YOLOv8 Ours (DCE-Net)

Figure 10. Visualization detection results of different methods on the SD-Dataset and the SQ-Dataset. The first column represents the labeled images, the second column represents the visualization of detection results based on the YOLOv8 model, and the third column represents the visualization of detection results based on DCE-Net.

Table 3. Comparison of detection performance with different methods.

Detection Method	mAP@0.5:0.95 (%)	mAP@0.5 (%)	R (%)	F1 (%)	Params (M)	FLOPs (G)	FPS ($f \cdot s^{-1}$)
Faster R-CNN	36.2	83.05	85.08	68.7	137.1	370.2	6
RetinaNet	33.4	77.95	75.89	73.9	37.97	170.1	13
YOLOv8m	33.9	81.8	81.2	80	3	8.1	277
YOLOv9m	34.5	79.4	73.9	75	20.02	76.5	54
YOLOv10m	33.4	76.9	73.2	73	16.5	63.5	65
YOLO11s	37.7	85.3	80	79	2.58	6.3	294
Ours (DCE-Net)	41.6	87.3	83.5	81	5.19	9.7	217

Table 3 presents a performance comparison of various object detection methods across multiple key metrics, including mAP@0.5:0.95, mAP@0.5, recall, F1_Score, parameters, FLOPs, and Frames Per Second (FPS). As shown in the table, the proposed DCE-Net framework achieves 41.6% and 87.3% for mAP@0.5:0.95 and mAP@0.5, respectively, significantly outperforming other comparative models such as Faster R-CNN, RetinaNet, and different versions of the YOLO series. Although Faster R-CNN has a slightly higher recall, its parameter count and computational cost are substantially higher than those of DCE-Net, indicating that DCE-Net maintains high precision while offering superior computational efficiency. Moreover, DCE-Net attains an inference speed of 217 $f \cdot s^{-1}$, underscoring its computational efficiency and suitability for latency-sensitive, real-time applications. YOLO11s exhibits the best performance in terms of parameter count and computational cost, but its mAP@0.5:0.95 and mAP@0.5 values are still lower than those of DCE-Ne. Although YOLO11s delivers a slightly higher throughput of 294 $f \cdot s^{-1}$, DCE-Net's superior accuracy combined with a competitive frame rate presents a more balanced trade-off for practical deployment scenarios. RetinaNet and other versions of the YOLO series show moderate performance across all metrics, particularly lagging behind DCE-Net in mAP@0.5:0.95 and recall. Considering both performance metrics and model complexity, DCE-Net demonstrates high practicality and expands value in underwater target detection tasks.

5. Conclusions

This paper proposes an improved method for sonar small-target detection, DCE-Net, based on YOLOv8, to tackle the challenge of small-target detection in underwater sonar images. The core advantage of DCE-Net lies in its multi-level feature fusion strategy, and it is the first to combine the DEAB with the CoordGate module for underwater sonar small-target detection. Simultaneously, to further enhance the feature representation capability, this paper designs a novel channel AM, namely MH-EMA, which further optimizes the feature fusion process by introducing the multi-head AM. The experimental results demonstrate that DCE-Net achieves significant performance improvements on the self-constructed sonar image dataset, SD-Dataset, with mAP@0.5 and mAP@0.5:0.95 reaching 87.3% and 41.6%, respectively, outperforming other comparative models. DCE-Net achieves high precision in detecting sonar small targets in complex underwater environments, and these results indicate that comprehensive consideration of image dehazing, feature extraction, and feature fusion can reduce false detections and improve detection accuracy. This research provides an effective solution for the field of underwater target detection, offering high practical value, but there remains scope for enhancing detection accuracy. Future work will center on further elevating the efficiency and exactitude of underwater target detection.

Author Contributions: Conceptualization, L.C.; methodology, L.C.; software, Z.M.; validation, Z.M., Q.H., Z.X., and M.Z.; formal analysis, L.C.; investigation, Z.M. and Q.H.; data curation, M.Z.; writing—original draft preparation, L.C.; writing—review and editing, L.C. and Z.M.; supervision, Q.H.; project administration, Z.M.; funding acquisition, Z.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under grants 62305123, 42176187, and 62401601.

Data Availability Statement: The datasets used in this paper involve sensitive information and cannot be provided externally in accordance with relevant confidentiality regulations.

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

References

1. Rioblanc, M. High productivity multi-sensor seabed Mapping Sonar for Marine Mineral Resources Exploration. In Proceedings of the 2013 IEEE International Underwater Technology Symposium (UT), Tokyo, Japan, 5–8 March 2013.
2. Character, L.; Ortiz JR, A.; Beach, T.; Luzzadder-Beach, S. Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar. *Remote Sens.* **2021**, *13*, 1759. [CrossRef]
3. Köhntopp, D.; Lehmann, B.; Kraus, D.; Birk, A. Classification and Localization of Naval Mines With Superellipse Active Contours. *IEEE J. Ocean. Eng.* **2019**, *44*, 767–782. [CrossRef]
4. Grothues, T.M.; Newhall, A.E.; Lynch, J.F.; Vogel, K.S.; Gawarkiewicz, G.G. High-frequency side-scan sonar fish reconnaissance by autonomous underwater vehicles. *Can. J. Fish. Aquat. Sci.* **2017**, *74*, 240–255. [CrossRef]
5. Wu, Q.; Liu, Y.; Li, Q.; Jin, S.; Li, F. The application of deep learning in computer vision. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017.
6. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [CrossRef] [PubMed]
7. Reddy, G.P.O. Digital Image Processing: Principles and Applications. In *Geospatial Technologies in Land Resources Mapping, Monitoring and Management*; Reddy, G.P.O., Singh, S.K., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 101–126.
8. Yin, J.-L.; Huang, Y.-C.; Chen, B.-H.; Ye, S.-Z. Color Transferred Convolutional Neural Networks for Image Dehazing. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 3957–3967. [CrossRef]
9. Su, J.; Xu, B.; Yin, H. A survey of deep learning approaches to image restoration. *Neurocomputing* **2022**, *487*, 46–65. [CrossRef]
10. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E.H. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef] [PubMed]
11. Liu, R.; Monfort, M.A.S.; Huang, J.Y.; Berg, A.S.; Hays, J. An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution. *arXiv* **2018**, arXiv:1807.03247v2.
12. Li, Z.; Xie, Z.; Duan, P.; Kang, X.; Li, S. Dual Spatial Attention Network for Underwater Object Detection with Sonar Imagery. *IEEE Sens. J.* **2024**, *24*, 6998–7008. [CrossRef]
13. Chen, Y.; Yuan, X.; Wang, J.; Wu, R.; Li, X.; Hou, Q.; Cheng, M.M. YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-time Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 4240–4252. [CrossRef]
14. Cao, Y.; Cui, X.D.; Gan, M.Y.; Wang, Y.X.; Yang, F.L.; Huang, Y. MAL-YOLO: A lightweight algorithm for target detection in side-scan sonar images based on multi-scale feature fusion and attention mechanism. *Int. J. Digit. Earth* **2024**, *17*, 2398050. [CrossRef]
15. Chen, Z.; He, Z.; Lu, Z.M. DEA-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention. *IEEE Trans. Image Process.* **2024**, *33*, 1002–1015. [CrossRef]
16. Howard, S.; Norreys, P.; Döpp, A. CoordGate: Efficiently Computing Spatially-Varying Convolutions in Convolutional Neural Networks. *arXiv* **2024**, arXiv:2401.04680. [CrossRef]
17. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *arXiv* **2023**, arXiv:2305.13563v2.
18. Celik, T.; Tjahjadi, T. A Novel Method for Sidescan Sonar Image Segmentation. *IEEE J. Ocean. Eng.* **2011**, *36*, 186–194. [CrossRef]
19. Saad, N.H.; Isa, N.A.M.; Saleh, H.M. Nonlinear Exposure Intensity Based Modification Histogram Equalization for Non-Uniform Illumination Image Enhancement. *IEEE Access* **2021**, *9*, 93033–93061. [CrossRef]
20. Yin, M.; Yang, J. ILR-Net: Low-light image enhancement network based on the combination of iterative learning mechanism and Retinex theory. *PLoS ONE* **2025**, *20*, e0314541. [CrossRef] [PubMed]
21. Li, H.; Xu, K. Innovative adaptive edge detection for noisy images using wavelet and Gaussian method. *Sci. Rep.* **2025**, *15*, 5838. [CrossRef] [PubMed]
22. Choudhary, R.R.; Jisnu, K.K.; Meena, G. Image DeHazing Using Deep Learning Techniques. *Procedia Comput. Sci.* **2020**, *167*, 1110–1119. [CrossRef]
23. Fu, H.; Ling, Z.; Sun, G.; Ren, J.; Zhang, A.; Zhang, L.; Jia, X. HyperDehazing: A hyperspectral image dehazing benchmark dataset and a deep learning model for haze removal. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 663–677. [CrossRef]

24. Wang, D.; Wang, Z. Research and Implementation of Image Dehazing Based on Deep Learning. In Proceedings of the 2022 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 23–25 September 2022.
25. Babu, G.H.; Odugu, V.K.; Venkatram, N.; Satish, B.; Revathi, K.; Rao, B.J. Development and performance evaluation of enhanced image dehazing method using deep learning networks. *J. Vis. Commun. Image Represent.* **2023**, *97*, 103976. [CrossRef]
26. Li, Z.; Zheng, C.; Shu, H.; Wu, S. Single Image Dehazing via Model-Based Deep-Learning. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022.
27. Vishnoi, R.; Goswami, P.K. A Comprehensive Review on Deep Learning based Image Dehazing Techniques. In Proceedings of the 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 16–17 December 2022.
28. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [CrossRef]
29. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [CrossRef]
30. Lin, K.; Wang, G.; Li, T.; Wu, Y.; Li, C.; Yang, Y.; Shen, H.T. Toward Generalized and Realistic Unpaired Image Dehazing via Region-Aware Physical Constraints. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *35*, 2753–2767. [CrossRef]
31. Wang, Y.; Yan, X.; Wang, F.L.; Xie, H.; Yang, W.; Zhang, X.-P.; Qin, J.; Wei, M. UCL-Dehaze: Toward Real-World Image Dehazing via Unsupervised Contrastive Learning. *IEEE Trans. Image Process.* **2024**, *33*, 1361–1374. [CrossRef]
32. Zheng, Y.; Su, J.; Zhang, S.; Tao, M.; Wang, L. Dehaze-TGGAN: Transformer-Guide Generative Adversarial Networks With Spatial-Spectrum Attention for Unpaired Remote Sensing Dehazing. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5634320. [CrossRef]
33. Engin, D.; Genç, A.; Kemal Ekenel, H. Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018.
34. Sahu, G.; Seal, A.; Yazidi, A.; Krejcar, O. A Dual-Channel Dehaze-Net for Single Image Dehazing in Visual Internet of Things Using PYNQ-Z2 Board. *IEEE Trans. Autom. Sci. Eng.* **2024**, *21*, 305–319. [CrossRef]
35. Wang, Y.K.; Fan, C.T. Single Image Defogging by Multiscale Depth Fusion. *IEEE Trans. Image Process.* **2014**, *23*, 4826–4837. [CrossRef] [PubMed]
36. Ullah, H.; Muhammad, K.; Irfan, M.; Anwar, S.; Sajjad, M.; Imran, A.S.; de Albuquerque, V.H.C. Light-DehazeNet: A Novel Lightweight CNN Architecture for Single Image Dehazing. *IEEE Trans. Image Process.* **2021**, *30*, 8968–8982. [CrossRef]
37. Sahu, G.; Seal, A.; Jaworek-Korjakowska, J.; Krejcar, O. Single Image Dehazing via Fusion of Multilevel Attention Network for Vision-Based Measurement Applications. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 4503415. [CrossRef]
38. Nie, J.; Wei, W.; Zhang, L.; Yuan, J.; Wang, Z.; Li, H. Contrastive Haze-Aware Learning for Dynamic Remote Sensing Image Dehazing. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5634311. [CrossRef]
39. 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.
40. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767v1.
41. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934v1.
42. Cao, X.; Ren, L.; Sun, C.Y. Dynamic Target Tracking Control of Autonomous Underwater Vehicle Based on Trajectory Prediction. *IEEE Trans. Cybern.* **2023**, *53*, 1968–1981. [CrossRef] [PubMed]
43. Lu, Y.; Zhang, L.; Xie, W. YOLO-compact: An Efficient YOLO Network for Single Category Real-time Object Detection. In Proceedings of the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020.
44. Li, X.; Cai, K. Method research on ship detection in remote sensing image based on Yolo algorithm. In Proceedings of the 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Xi'an, China, 14–16 August 2020.
45. Yang, N.; Li, G.; Wang, S.; Wei, Z.; Ren, H.; Zhang, X.; Pei, Y. SS-YOLO: A Lightweight Deep Learning Model Focused on Side-Scan Sonar Target Detection. *J. Mar. Sci. Eng.* **2025**, *13*, 66. [CrossRef]
46. Ge, J.; Zhang, B.; Wang, C.; Xu, C.; Tian, Z.; Xu, L. Azimuth-Sensitive Object Detection in Sar Images Using Improved Yolo V5 Model. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022.
47. Liu, C.; Tao, Y.; Liang, J.; Li, K.; Chen, Y. Object Detection Based on YOLO Network. In Proceedings of the 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 14–16 December 2018.
48. Wang, X.; Jiang, X.; Xia, Z.; Feng, X. Underwater Object Detection Based on Enhanced YOLO. In Proceedings of the 2022 International Conference on Image Processing and Media Computing (ICIPMC), Xi'an, China, 27–29 May 2022.
49. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

50. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
51. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
52. Wang, Z.; Guo, J.; Zeng, L.; Zhang, C.; Wang, B. MLFFNet: Multilevel Feature Fusion Network for Object Detection in Sonar Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5119119. [CrossRef]
53. Li, H.; Miao, S.; Feng, R. DG-FPN: Learning Dynamic Feature Fusion Based on Graph Convolution Network For Object Detection. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020.
54. Tong, X.; Zuo, Z.; Sun, B.; Wei, J. Novel Feature Fusion for Infrared Small Target Detection Feature Pyramid Networks. In Proceedings of the 2021 IEEE 9th International Conference on Information, Communication and Networks (ICICN), Xi'an, China, 25–28 November 2021.
55. Zhang, S.; Zhang, X.; Zhang, A.; Fu, H.; Cheng, J.; Huang, H.; Sun, G.; Zhang, L.; Yao, Y. Fusion Of Low-And High-Level Features For Uav Hyperspectral Image Classification. In Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 September 2019.
56. Chen, J.; Mai, H.; Luo, L.; Chen, X.; Wu, K. Effective Feature Fusion Network in BIFPN for Small Object Detection. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021.
57. Zhang, Y.; Cheng, J.; Bai, H.; Wang, Q.; Liang, X. Multilevel Feature Fusion and Attention Network for High-Resolution Remote Sensing Image Semantic Labeling. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6512305. [CrossRef]
58. Wang, J.; Ren, Y.; Wei, S. Synthetic Aperture Radar Images Target Detection and Recognition with Multiscale Feature Extraction and Fusion Based on Convolutional Neural Networks. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019.
59. Hu, Z.; Kong, Q.; Liao, Q. Multi-Level Adaptive Attention Fusion Network for Infrared and Visible Image Fusion. *IEEE Signal Process. Lett.* **2025**, *32*, 366–370. [CrossRef]
60. Liang, M.; Jiao, L.; Yang, S.; Liu, F.; Hou, B.; Chen, H. Deep Multiscale Spectral-Spatial Feature Fusion for Hyperspectral Images Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2911–2924. [CrossRef]
61. Yu, L.; Hu, H.; Zhong, Z.; Wu, H.; Deng, Q. GLF-Net: A Target Detection Method Based on Global and Local Multiscale Feature Fusion of Remote Sensing Aircraft Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4021505. [CrossRef]
62. Cao, Z.; Chen, K.; Chen, J.; Chen, Z.; Zhang, M. CACS-YOLO: A Lightweight Model for Insulator Defect Detection Based on Improved YOLOv8m. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3530710. [CrossRef]
63. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
64. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]

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.