

Article

Research on a Lightweight Detection Method for Underwater Diseased Corals

Mingqi Li and Ming Chen * 

Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; lmq18029510200@163.com

* Correspondence: mchen@shou.edu.cn

Abstract

In underwater detection tasks involving bleached corals, band disease corals, and white pox disease corals, several challenges persist, including high morphological variability, difficulty in identifying small pathological regions, interference from complex underwater environments, and constraints imposed by underwater hardware. To address these issues, a lightweight underwater diseased coral target detection method, termed CD-YOLO, is proposed. Specifically, (1) a lightweight network named CDShuffleNet is constructed to replace the YOLO11 backbone, aiming to reduce model complexity while preserving detection performance; (2) a SPDConv downsampling convolution module is introduced to reduce the loss of fine-grained coral detail information during the downsampling process; and (3) attention mechanisms are incorporated through an engineering-oriented integration of EMA into the C2PSA module and the adoption of SENetV2, in order to enhance the representation of color and shape features of pathological regions and suppress interference from complex underwater environments. Experimental results demonstrate that the proposed improvements yield consistent gains in both model lightweighting and detection performance under the adopted evaluation settings. Specifically, the number of parameters, computational cost, and model size are reduced by 20.6%, 21.9%, and 18.9%, respectively, while mAP increases by 4.3 percentage points. Comparative experiments further show that the proposed method achieves a markedly higher mAP than several other state-of-the-art models. In addition, experiments conducted on the BHD Coral dataset provide preliminary evidence of cross-dataset adaptability of the proposed model. Overall, this study presents a task-oriented and application-driven improvement, demonstrating that the effective integration of lightweight components can achieve a favorable balance between model efficiency and detection performance in underwater diseased coral detection tasks.

Keywords: coral; YOLO11; ShuffleNet; deep learning; object detection



Academic Editor: Andrea Prati

Received: 20 January 2026

Revised: 3 February 2026

Accepted: 3 February 2026

Published: 5 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Coral reefs are among the most important structural components of marine ecosystems. Although they cover less than 1% of the ocean's surface area, they provide habitats and food resources for approximately 25% of marine organisms [1,2]. Common forms of coral abnormalities include coral bleaching, band disease, and white pox disease [3]. In recent years, the occurrence of abnormal coral phenomena has shown an increasing global trend, with coral bleaching becoming particularly severe, posing a major challenge to coral reef conservation both domestically and internationally [4,5]. If such abnormal coral conditions

persist over extended periods, they can lead to large-scale coral mortality, threaten coral-dependent organisms and food webs, and ultimately undermine the stability of marine ecosystems. Therefore, timely detection of corals and the implementation of appropriate artificial intervention measures based on the condition of abnormal corals can improve coral survival rates and help maintain marine ecological balance [6,7].

To address the aforementioned challenges, convolutional neural network (CNN)-based object detection methods have become the mainstream approach in underwater coral-related tasks. These methods can be broadly categorized into two types: two-stage algorithms and one-stage algorithms. Representative two-stage algorithms include R-CNN (Region-Based CNN) [8] and Faster R-CNN [9], while typical one-stage algorithms include YOLO (You Only Look Once) [10] and EfficientDet [11]. These approaches have been widely applied to underwater object detection tasks, including those related to underwater coral detection. Xin G. et al. employed an improved convolutional neural network architecture, EfficientNet, for the detection of bleached corals [12]. González-Rivero et al. utilized VGG-16 in combination with a support vector machine (SVM) to detect corals and benthic coral reef communities [13]. Wang Lan et al. applied an improved deep data augmentation method, DeepSMOTE, together with transfer learning for the detection of reef-building corals [14]. Hua Mingzhu adopted an improved YOLOv5 model to perform classification and detection of corals, reef fishes, and starfish [15]. Although these studies have achieved notable progress in coral-related detection tasks, existing methods still suffer from insufficient lightweight design and limited robustness when applied to diseased coral detection. In this work, robustness primarily refers to the model's ability to maintain stable detection performance by suppressing background interference and enhancing pathological feature representation under varying underwater imaging conditions.

Corals are characterized by high species diversity, complex and variable morphologies, and often dense spatial distributions. In addition, the widespread presence of interference factors such as fish further increases the difficulty of diseased coral detection. Moreover, due to underwater optical effects such as light attenuation and scattering, coral images captured underwater may suffer from severe color distortion and blurring. Factors such as water currents can also cause variations in target posture, further complicating coral object detection [16]. On the other hand, operations in underwater environments are typically constrained by limitations in computational capability and storage capacity of the deployed equipment [17]. Consequently, existing object detection models generally exhibit the following shortcomings: Methods with higher detection accuracy usually suffer from high computational complexity and large model sizes, making them difficult to deploy in underwater task environments; in contrast, methods with smaller model sizes and lower computational complexity that are easier to deploy often exhibit relatively lower detection accuracy and robustness. Therefore, it is necessary to develop a lightweight underwater diseased coral object detection algorithm that achieves a better balance between model efficiency, detection accuracy, and robustness.

2. Method

Underwater diseased coral detection faces practical challenges such as complex illumination conditions, low contrast, background interference, and limited computational resources in real deployment scenarios. Therefore, the proposed method follows a task-oriented design strategy, aiming to balance detection accuracy and computational efficiency.

To achieve more accurate, lightweight, and easily deployable underwater diseased coral object detection, this study proposes multiple improvements to the YOLO11 model, resulting in the CD-YOLO architecture. The overall network structure is illustrated in Figure 1.

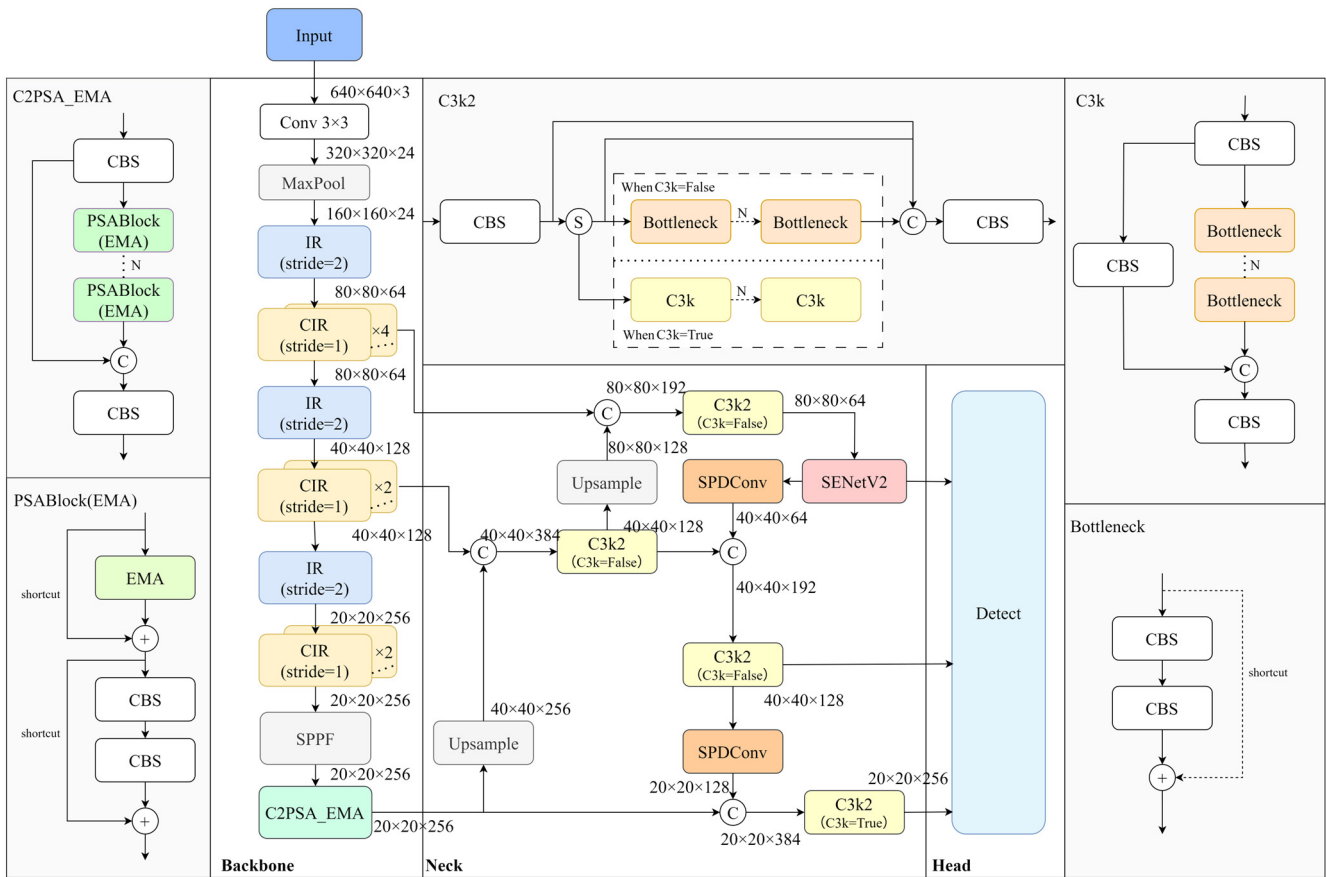


Figure 1. CD-YOLO Model Structure. IR denotes the original core module of ShuffleNetV2. CIR represents the improved core module designed for CDShuffleNet. The value in parentheses indicates the stride parameter of the corresponding module, and the ellipsis multiplied by a number denotes that the IR or CIR module is stacked the specified number of times at that position. The detailed structures of the IR and CIR modules are shown in Figure 2. A white circle containing a “+” symbol represents a residual connection operation, a “C” symbol denotes a concatenation operation, and an “S” symbol indicates a split operation. The CBS module consists of a convolution (Conv), batch normalization (BN), and a SiLU activation function. The C2PSA_EMA module is an improved version of C2PSA. SPDConv is a convolutional module, while EMA and SENetV2 are both attention modules.

The main architectural design choices of CD-YOLO are summarized below: (1) The core unit of ShuffleNetV2 (ShuffleNet Version 2) is improved by adopting a “dimension reduction followed by dimension expansion” strategy combined with residual connections. Based on this design, a task-oriented lightweight network named CDShuffleNet (Coral Disease–ShuffleNet) is constructed by adapting and reorganizing existing lightweight design principles for underwater diseased coral detection, which replaces the backbone of YOLO11 to achieve model lightweighting while simultaneously improving detection performance; (2) the downsampling convolution module in the Neck is optimized to SPDConv (Space-to-Depth Convolution), which further enhances detection accuracy while maintaining a lightweight model design; and (3) attention mechanisms are integrated, including an improved C2PSA module incorporating the EMA (Efficient Multi-scale Attention) mechanism, as well as the fusion of the SENetV2 (Squeeze-and-Excitation Network Version 2) attention mechanism in the Neck. These designs further promote model lightweighting, strengthen feature representation capability, and improve the overall robustness of the model.

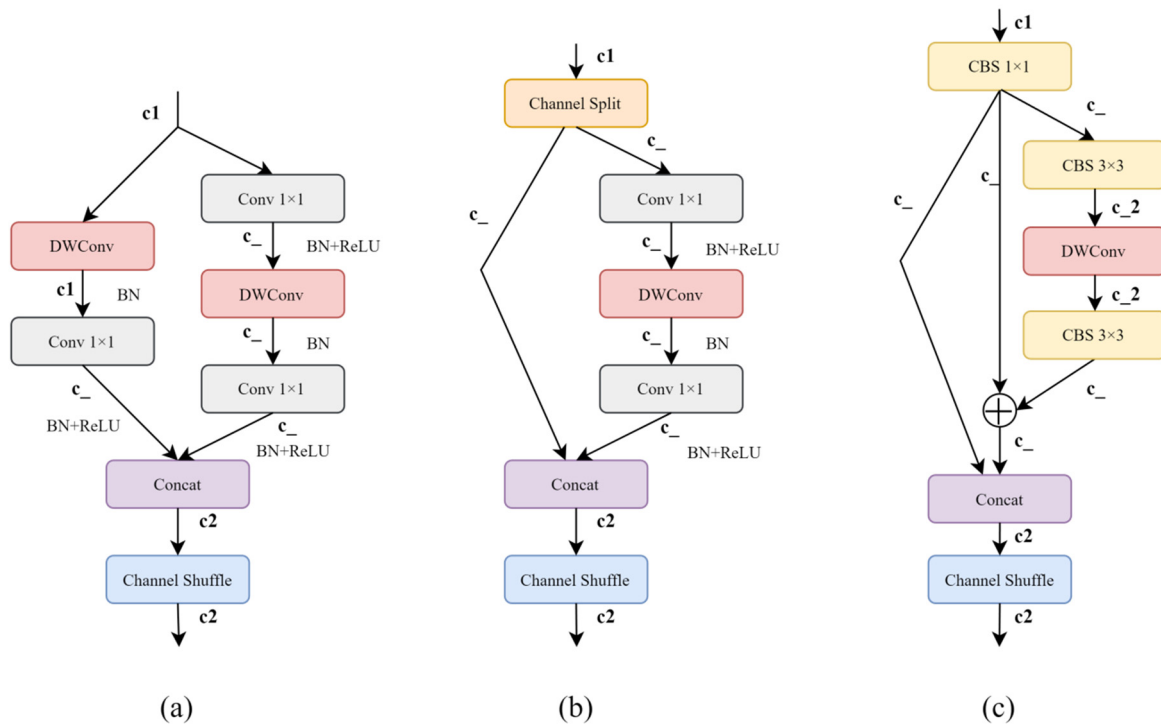


Figure 2. Original Core Module Structure of ShuffleNetV2 and the Improved Core Module Structure. (a) shows the core module with a stride value of 2, (b) presents the original core module with a stride value of 1, and (c) depicts the improved core module with a stride value of 1. A white circle containing a “+” symbol represents a residual connection, and the CBS module consists of Conv + BN + SiLU. And c_1 denotes the number of input channels, c_2 denotes the number of output channels, c_- represents half of c_2 , and c_{-2} represents half of c_- .

2.1. YOLO11 Model

YOLO11 is one of the latest versions in the YOLO series. It adopts improved backbone, Neck, and Head architectures, which enhance feature extraction capability to achieve more accurate object detection and high performance in complex tasks. The model provides faster processing speed while maintaining a favorable balance between accuracy and efficiency, and it can be deployed in various environments, including edge devices, demonstrating high flexibility [18].

Owing to its excellent trade-off between detection accuracy and inference speed, YOLO11 is often selected as a baseline and further improved for resource-constrained object detection tasks [19]. Compared with YOLO12 in the YOLO series and various Transformer-based models, YOLO11 has lower computational complexity and smaller model size. Therefore, this study selects YOLO11 as the baseline model for improvement. By enhancing the YOLO11 model, it is possible to not only improve the overall detection accuracy in underwater diseased coral object detection tasks, but also further reduce the number of model parameters, computational cost, and model size, thereby making the model more suitable for deployment.

2.2. Backbone Replacement with CDShuffleNet

Due to the limitations of underwater hardware conditions in diseased coral detection tasks, the detection model is required to maintain a balance between computational efficiency and feature representation capability. Therefore, a lightweight backbone architecture is adopted in this study.

This study improves the core unit module of the lightweight ShuffleNetV2 network by drawing on the characteristics of ShuffleNetV2 and multiple network design principles.

The improved core unit modules are then used to construct a new network, CDShuffleNet, in which the number of unit modules at each stage is reduced compared with the original network. As a result, CDShuffleNet is lighter than the original YOLO11 backbone while maintaining comparable detection performance, making it suitable for object detection tasks in constrained underwater environments. The details are described as follows.

ShuffleNetV2 [20] is a lightweight and efficient deep neural network specifically designed for computation-constrained environments such as mobile devices. Based on ShuffleNetV1, it emphasizes that, in addition to FLOPs, the relationship between the number of multiply–accumulate operations and actual inference latency should also be considered. ShuffleNetV2 adopts a more uniform channel allocation strategy, reduces memory access cost, and introduces new computational units to improve parallel computing efficiency. Its main characteristics include: splitting the input channels into two branches, where one branch undergoes lightweight computation while the other is directly bypassed; removing the group convolution (GConv) used in ShuffleNetV1; employing channel shuffle operations to ensure effective information fusion; and reducing the proportion of point-wise operations.

Residual connections and Bottleneck modules were first introduced in ResNet [21]. In traditional neural networks, each layer typically learns a mapping function $H(x)$, whereas residual connections indirectly learn this mapping by optimizing a residual function $F(x)$. This mechanism effectively alleviates gradient vanishing and gradient explosion problems as network depth increases. The Bottleneck structure is designed to address the growth in computational cost and parameter count caused by deeper networks. In YOLO11, the Bottleneck module consists of a CBS module that first reduces the channel dimension, followed by another CBS module that restores the channel dimension, with an optional residual connection. This “dimension reduction followed by dimension expansion” design reduces computational cost while maintaining model representational capacity.

The original core module of ShuffleNetV2 and the improved core module proposed in this study are illustrated in Figure 2. A stride value of 1 is used for feature extraction, while a stride value of 2 is used for downsampling. In this work, the modules with a stride value of 2 are retained to ensure lightweight downsampling. The modules with a stride value of 1 are redesigned as follows.

First, the channel splitting operation is removed, and a CBS module with a convolution kernel size of 1 is employed to compress the channel dimension. The compressed features are then divided into three branches, referred to as Branch 1, Branch 2, and Branch 3. Next, Branch 3 is sequentially processed by a CBS module with a kernel size of 3, a depthwise convolution (DWConv), and another CBS module with a kernel size of 3. In this process, the first CBS module performs channel compression (“dimension reduction”), while the second CBS module restores the channel dimension (“dimension expansion”). Subsequently, the processed Branch 3 is added element-wise to Branch 2 to form a residual connection. The output of this residual connection is then concatenated with Branch 1, followed by a channel shuffle operation to obtain the final output of the module.

Compared with the original core module, the improved module removes the original Conv 1×1 + BN + ReLU combination and replaces it with CBS modules, with only the first CBS module using a 1×1 convolution kernel. By increasing the use of 3×3 convolutions and adopting the SiLU activation function, the feature learning capability of the module is effectively enhanced, although this also increases computational cost. Therefore, in contrast to the original module, the feature extraction branch is further compressed to a lower channel dimension before applying the DWConv from the original module and subsequently restoring the channel dimension. This design reduces computational cost while preserving sufficient feature representation capability, such as learning shape and

color information. Moreover, since ShuffleNetV2 stacks a large number of core modules, deep networks are prone to gradient vanishing. To alleviate this issue, a residual connection is introduced before channel concatenation in the improved core module.

In addition, because the improved core modules exhibit stronger feature extraction capability, the ShuffleNetV2 network architecture is further adjusted by reducing the number of stacked core modules at each stage, thereby decreasing the overall network complexity and computational cost. The original ShuffleNetV2 architecture and the proposed CDSHuffleNet architecture are illustrated in Figure 3.

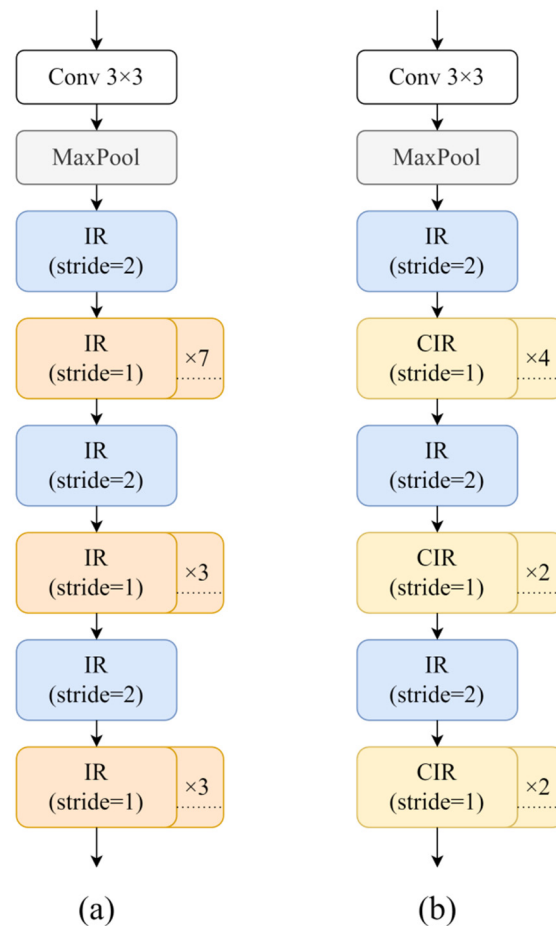


Figure 3. Original ShuffleNetV2 Network Structure and CDSHuffleNet Network Structure. (a) shows the original ShuffleNetV2 network and (b) presents the CDSHuffleNet network. IR denotes the original core module of ShuffleNetV2. CIR represents the improved core module designed for CDSHuffleNet. The value in parentheses indicates the stride parameter of the corresponding module. The ellipsis and the multiplied number displayed on the IR or CIR modules indicate the number of times that the corresponding module is stacked at that position in the network.

CDSHuffleNet is employed to replace the backbone network of YOLO11, aiming to reduce model complexity while preserving essential feature extraction capability under underwater conditions. Owing to uneven underwater illumination, light refraction and scattering can cause brightness variations, while light attenuation often leads to color shifts, resulting in coral images that are sometimes blurred and thus hinder effective feature extraction. Moreover, abnormal coral regions vary in size and may exhibit low contrast with healthy corals, making them difficult to distinguish. Bleached corals mainly differ from healthy corals in terms of color, whereas other types of diseased corals differ primarily in fine-grained details, such as patches and stripes.

By adopting the improved core unit modules, CDShuffleNet is able to learn color information as well as detailed features such as patches and stripes on corals more effectively than the original ShuffleNet network. This enhancement strengthens the overall feature extraction capability of the model, enabling it to maintain favorable detection performance while achieving lightweight design, and making it better suited to meet the deployment and detection requirements in underwater environments.

2.3. Incorporation of SPDConv

Underwater coral images often exhibit insufficient resolution, reduced contrast, and significant noise due to limitations of imaging equipment as well as light scattering and absorption effects in water. When coral targets are small in scale, sparsely distributed, or located at a considerable distance from the imaging device, the model's ability to perceive fine-grained features and small objects is severely challenged.

Existing studies have demonstrated that, in scenarios with significant background interference or limited data quality, preserving and effectively processing informative features is critical to the performance of deep learning models. For example, Selvan et al. employed discrete wavelet transform (DWT) to remove noise from sensor signals and achieved an accuracy of 99.58% in a fetal health classification task using deep learning models [22]. This result further indicates that strengthening information preservation in complex environments has general significance for improving model performance and is also worthy of attention in underwater object detection tasks.

To address this issue, this study improves the downsampling module, as described below. SPDConv [23] is a downsampling convolution module. In conventional convolutional neural networks, strided convolutions and pooling operations are widely used for feature map downsampling to enlarge the receptive field and reduce computational cost. However, when the input image resolution is low or the target objects are small, early-stage downsampling using these methods discards a substantial amount of fine-grained information, leading to significant performance degradation in small-object detection and low-resolution classification tasks. SPDConv splits the input feature map into multiple sub-feature maps using a fixed stride and then concatenates them along the channel dimension to achieve downsampling. Although the spatial resolution is reduced, the number of channels is increased, preserving more feature information. A convolution with a stride of 1 is then applied to reduce the channel dimension. In this way, fine-grained information loss that could lead to insufficient feature learning is avoided, while the computational overhead of the downsampling module is not increased.

In YOLO11, the Neck still employs several conventional downsampling convolution modules. In this work, these modules are replaced with SPDConv. After this replacement, the model is able to maintain sufficient feature learning of coral details under low-resolution conditions or when targets are small. Consequently, the proposed improvement enhances the model's capability for underwater diseased coral detection.

2.4. Fusion of Attention Mechanisms

After the aforementioned improvements, the model achieves a certain degree of lightweight design while maintaining relatively high detection accuracy. However, in practical scenarios, underwater coral detection remains highly challenging due to the presence of small and scattered pathological regions, such as localized bleaching and white pox disease occurring among numerous healthy corals. In addition, white pathological areas are easily confused with white sandy seabeds, and interference from water surface reflections, illumination variations, and marine organisms further increases the complexity of the detection task. As a result, conventional models often exhibit limited robustness

in terms of stable feature discrimination and background suppression under complex underwater environments. To address these issues, this study integrates two attention mechanisms to further lighten the model while enhancing the representation of pathological region features, suppressing environmental interference, and improving overall robustness. The details are as follows.

Efficient Multi-scale Attention (EMA) [24] is an efficient multi-scale attention module specifically designed for computer vision tasks, which is able to reduce the number of parameters and computational cost while preserving the key information of each channel. EMA enhances feature processing capability by reorganizing channel and batch dimensions, captures pixel-level relationships through cross-dimensional interactions, and improves feature representation via global information encoding and channel weight calibration. These designs enable EMA to achieve a favorable balance between the number of parameters, computational cost, and detection performance, thereby facilitating further model lightweighting. SENetV2 [25] is an improved version of the Squeeze-and-Excitation Network (SENet), whose core component is the Squeeze Aggregated Excitation (SaE) module. By combining the characteristics of SENet and ResNeXt architectures, SaE adopts multi-branch fully connected layers for feature compression followed by scaling. This design enhances global information aggregation and feature representation capability, allowing SENetV2 to learn more complex channel relationships with negligible additional computational cost.

Accordingly, this study integrates the EMA mechanism into the C2PSA module to construct the C2PSA_EMA module. The C2PSA module is an innovative component of YOLO11 and serves as the final processing module in the Backbone. In this work, Attention modules within all PSABlock units of the original C2PSA module are replaced with EMA. EMA modules are more lightweight than the original Attention modules. By leveraging EMA's cross-dimensional interaction and global information encoding capabilities, the model can learn the relationships between different regions of an image, enabling more effective discrimination between targets and background across diverse environments.

Furthermore, the SENetV2 mechanism is introduced into the Neck of YOLO11, aiming to strengthen channel-wise feature discrimination for small and visually ambiguous coral targets. By learning complex channel relationships and aggregating global information, SENetV2 enhances the model's sensitivity to pathological features in terms of color (e.g., white regions) and shape (e.g., band-like or spot-like patterns). Its placement in the small-object detection layer enables more effective handling of small and dispersed white pox disease regions or localized bleaching mixed within healthy corals, thereby improving the overall robustness of the model.

Since the above attention mechanism modules are inherently lightweight, the integration of the EMA mechanism replaces the original attention module, while the SENetV2 mechanism is only added to a single detection layer, resulting in a limited increase in parameters and computational cost. Therefore, the incorporation of these two attention mechanisms allows the improved model to maintain its lightweight characteristics.

3. Experimental Design and Results Analysis

3.1. Experimental Design

3.1.1. Experimental Parameters and Experimental Environment

To ensure experimental reproducibility, all training processes were conducted using a unified set of hyperparameter configurations. The initial learning rate was set to 0.01, the momentum coefficient was set to 0.937, and the weight decay coefficient was set to 0.0005. The stochastic gradient descent (SGD) optimizer was adopted. All experiments in this study were trained for a total of 150 epochs.

The experimental environment for all conducted experiments is as follows: the operating system was Windows 11, the programming language was Python 3.12, the CPU was an AMD Ryzen 7 7745HX with Radeon Graphics, and the GPU was an NVIDIA GeForce RTX 4070 Laptop GPU with 8188 MiB of memory. The deep learning framework used was PyTorch 2.4.1, and GPU acceleration was supported by CUDA 12.0 and cuDNN 9.0.1.

3.1.2. Dataset and Preprocessing

A publicly available dataset named “coral” (Project ID: coral-z3riv-g4bwh) from the Roboflow platform was utilized in this study. The dead coral category, which contained a small number of samples and exhibited poor image quality, along with its related data, was removed. The resulting dataset consisted of 913 underwater coral images in JPG format, covering four categories: band disease corals, white pox disease corals, bleached corals, and healthy corals. The dataset includes challenging detection scenarios such as poor illumination, dense small targets, mixed multi-object scenes, and low image resolution, thereby ensuring the robustness of the proposed model under diverse real-world conditions.

Subsequently, the online annotation tool provided by Roboflow was used for data labeling. For categories retained in this study, the original labels from the source dataset were preserved, and the bounding box locations remained unchanged. For categories not used in this study, the corresponding labels were removed. Images with no remaining annotations after label removal were directly excluded from the dataset. After annotation, TXT-format annotation files were generated for experimental use.

Due to the limited number of coral samples available, in order to improve the model’s recognition accuracy and simultaneously simulate various complex underwater environments and detection challenges, such as different lighting conditions, the dataset was subjected to the following preprocessing and data augmentation operations. Specifically, all images were uniformly resized to 640×640 pixels using stretching; horizontal flipping was applied with a probability of 50%; random cropping was performed within a range of 0–20%; random rotation was applied within a range of -15° to $+15^\circ$; and image brightness was randomly adjusted within -12% to $+12\%$ of the original brightness. After these operations, the total number of samples was expanded to 2213 JPG images, including 1950 images for training, 182 images for validation, and 81 images for testing. Examples of the dataset and augmented samples are shown in Figure 4.

In addition, to verify the generalization capability of the proposed model, the publicly available BHD Coral dataset [26] was introduced. This dataset is a high-quality object detection dataset containing bleached corals as well as healthy and dead corals, and it covers a wide variety of underwater coral scenes. The dataset consists of 1275 underwater coral images in JPG format. “generalization” in this study is used in a restricted sense, referring to the model’s adaptability to datasets with similar imaging conditions and task characteristics, rather than broad cross-domain generalization across substantially different underwater environments.

Given that the data volume remains relatively limited, data augmentation was applied to this dataset using the following strategies: random adjustment of image brightness within a range of -15% to $+15\%$ of the original brightness, followed by the addition of noise to 0.1% of randomly selected pixels. After data augmentation, the dataset was expanded to 3155 JPG images, including 2820 images for training, 234 images for validation, and 101 images for testing. This augmented dataset effectively validates the generalization ability and robustness of the improved model for underwater coral detection tasks.

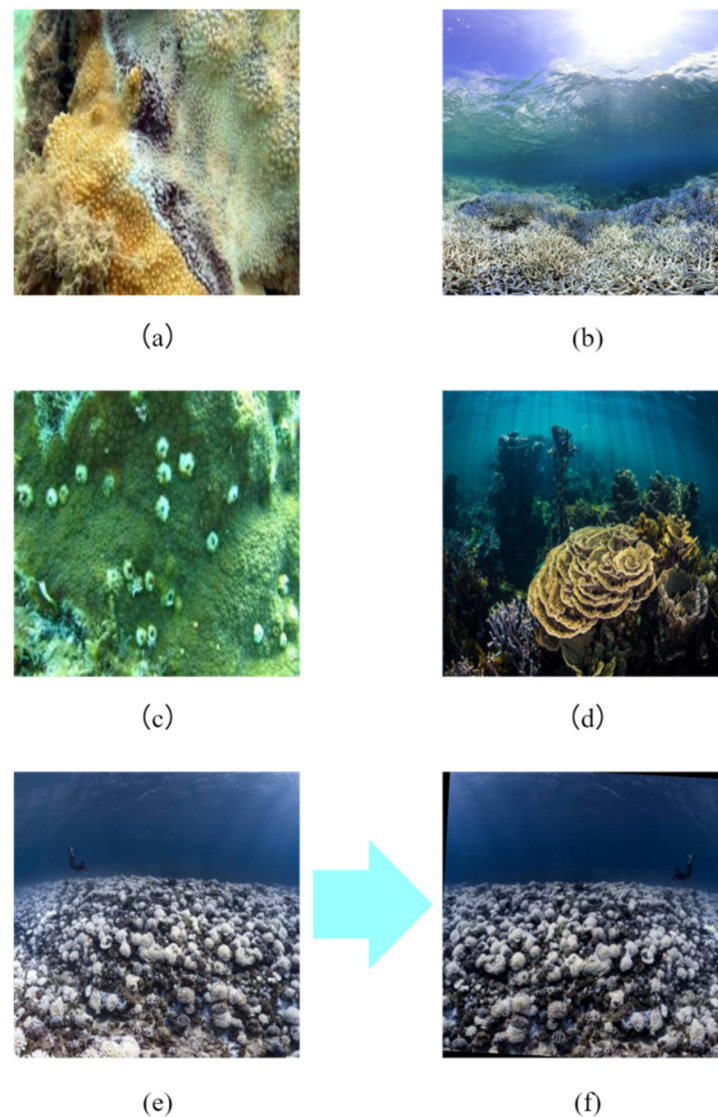


Figure 4. Dataset and Data Augmentation Example. (a) Band disease corals. (b) Bleached corals. (c) White pox disease corals. (d) Healthy corals. (e) An original image. (f) The augmented image after applying flipping, rotation, cropping, and brightness adjustment.

It should be noted that the size of the original dataset is relatively limited. Although data augmentation was employed to increase sample diversity and improve training stability, the augmented data cannot fully replace large-scale real-world underwater observations. Therefore, the experimental results reported in this study should be interpreted as indicative performance trends under the given dataset, rather than definitive conclusions applicable to all underwater coral detection scenarios.

3.1.3. Model Evaluation Metrics

To objectively, accurately, and comprehensively evaluate the performance of the proposed CD-YOLO model and the original YOLO11 model before and after improvements, as well as to assess the overall detection performance of different models in ablation and comparative experiments, this study adopts the following object detection evaluation metrics: Precision (P), recall (R), average precision (AP), mean average precision (mAP), number of model parameters (Params), floating-point operations (FLOPs), and model size (Size).

Among these metrics, average precision (AP) is a comprehensive indicator that jointly considers precision and recall. Precision (P) denotes the proportion of samples predicted

as positive that are actually positive, while recall (R) represents the proportion of actual positive samples that are correctly identified. The mean average precision (mAP) is defined as the mean of the average precision values across all categories. In underwater coral disease detection tasks, the target regions are often small, irregular in shape, and subject to annotation uncertainty caused by occlusion, illumination variation, and image degradation. Under such conditions, mAP computed at a fixed IoU threshold is commonly adopted in related studies as a stable and interpretable evaluation metric. Therefore, mAP is selected as the primary metric to facilitate fair comparison under consistent evaluation settings.

In this study, “lightweight” specifically refers to reductions in model scale and computational complexity, which are quantitatively measured by the number of model parameters, floating-point operations, and model size. These metrics are adopted to reflect the feasibility of deploying the model under resource-constrained underwater hardware conditions. Meanwhile, average precision and mean average precision are used to evaluate the overall detection accuracy of the model.

3.2. Experimental Analysis

Given the limited scale of the available datasets, the following experimental analysis focuses on comparative performance under identical training and evaluation settings, rather than absolute performance claims.

3.2.1. Results of the Improved Model

To verify the effectiveness of the proposed lightweight CD-YOLO model, all experiments in this study adopt YOLO11n, the most lightweight variant of YOLO11, as the representative baseline. The comparative results between the CD-YOLO model and the YOLO11n model are presented in Table 1.

Table 1. Comparison of the Model Before and After Improvement.

Model	Params	FLOPs/G	Size/MB	Precision/%	Recall/%	AP/%				mAP/%
						Band Disease	Bleached Disease	Healthy Coral	White Pox Disease	
YOLO11n	2,590,620	6.4	5.23	82.5	61.3	86.7	76.1	60.7	44.9	67.1
CD-YOLO	2,056,252	5.0	4.24	87.6	63.3	95.0	76.4	67.5	46.7	71.4

As shown by the experimental results in Table 1, compared with the baseline model, the improved model achieves reductions of 20.6%, 21.9%, and 18.9% in the number of parameters, computational complexity, and model size, respectively. These reductions indicate that the improved model effectively lowers computational cost and storage overhead, making it more feasible under resource-constrained conditions and more suitable for deployment in underwater environments.

In addition, compared with the baseline model, both precision and recall are improved, indicating a reduction in false-positive and false-negative rates. The mean Average Precision of the improved model increases by 4.3 percentage points, and the average precision for each category is also enhanced, with particularly notable improvements observed for band disease and healthy corals. These results further demonstrate that the overall detection performance of the improved model has been effectively enhanced.

3.2.2. Ablation Experiments

To verify the effectiveness of each proposed improvement strategy, ablation experiments were conducted using the YOLO11n model as the baseline. The improved modules

were gradually introduced, and the experimental results before and after the improvements were compared for evaluation. All models, both before and after improvement, were trained using the same dataset and identical training procedures. The design and results of the ablation experiments are presented in Tables 2 and 3, respectively.

Table 2. Ablation Experiments Design.

Model	Improvement Strategy			
	CDSuffleNet	SPDConv	C2PSA_EMA	SENetV2
YOLO11n				
YOLO11-CDS	✓			
YOLO11-SPD		✓		
YOLO11-SS	✓	✓		
YOLO11-SSE	✓	✓	✓	
CD-YOLO	✓	✓	✓	✓

Table 3. Ablation Experiments Results.

Model	Params	FLOPs/G	Size/MB	mAP/%
YOLO11n	2,590,620	6.4	5.23	67.1
YOLO11-CDS	2,207,260	5.2	4.53	69.9
YOLO11-SPD	2,488,220	6.3	5.03	70.1
YOLO11-SS	2,104,860	5.1	4.33	69.9
YOLO11-SSE	2,054,204	5.0	4.23	70.5
CD-YOLO	2,056,252	5.0	4.24	71.4

In the tables, YOLO11n denotes the original YOLO11 model with the smallest number of parameters; YOLO11-CDS refers to the model in which the backbone of YOLO11n is replaced by CDSuffleNet, obtained by improving ShuffleNetV2; YOLO11-SPD represents the model in which the downsampling convolution in the Neck layer of YOLO11n is replaced with SPDConv; YOLO11-SS is the model that simultaneously applies the above two improvements; and CD-YOLO, the final model proposed in this paper, is constructed by further integrating attention mechanisms (C2PSA_EMA and SENetV2) on the basis of YOLO11-SS.

The ablation experiment results indicate that, compared with the YOLO11n model, YOLO11-CDS achieves reductions in the number of parameters, computational cost, and model size after improving the backbone feature extraction network with the proposed CDSuffleNet, while simultaneously improving mAP. This demonstrates that the model has initially achieved superior lightweight characteristics and detection performance in the underwater diseased coral detection task.

In contrast, YOLO11-SPD replaces the conventional downsampling convolution module with SPDConv, resulting in an improvement in mAP while maintaining a similar model scale and computational cost, indicating that this modification has a significant effect on enhancing detection performance. By combining these two improvements, the YOLO11-SS model is obtained, which not only becomes more lightweight than YOLO11n but also exhibits a substantial overall improvement in detection performance.

After replacing C2PSA with C2PSA_EMA in YOLO11-SS, the resulting model is referred to as YOLO11-SSE. Experimental results show that not only the number of parameters, computational cost, and model size are reduced, but the mAP is also improved, demonstrating the effectiveness of using EMA to replace the attention module in C2PSA.

Furthermore, by integrating the SENetV2 attention mechanism into YOLO11-SSE, the CD-YOLO model is obtained. The mAP is further improved to 71.4%, while the

degree of lightweight design remains almost unchanged. This indicates the effectiveness of SENetV2 in small-object detection layers and ultimately confirms the superior performance of CD-YOLO on the experimental dataset used in this study.

3.2.3. Comparative Experiments

To further objectively and accurately evaluate the performance of the proposed CD-YOLO model, several other lightweight models were selected for comparison, including Transformer-based models RT-DETR-ResNet50 [27] and RT-DETR-HGNet [27]; other mainstream or recent models in the YOLO series, namely YOLOv8n [28], YOLOv10n [29], and YOLO12 [30]; as well as other lightweight models ModelA [31], ModelB [32], ModelC [33], and YOLO-Sh. Among them, ModelA, ModelB, and ModelC are lightweight models adopted in related studies, while YOLO-Sh is a model obtained by replacing the backbone network of YOLO11n with the original ShuffleNetV2. All models were trained using the same dataset and identical training strategies. The comparative experimental results of these models are presented in Table 4.

Table 4. Comparison Experiments Results.

Model	Params	FLOPs/G	Size/MB	mAP/%
RT-DETR-ResNet50	42,768,952	130.5	82.06	63.0
RT-DETR-HGNet	32,814,296	108.0	63.12	64.6
YOLOv8n	2,690,988	6.9	5.37	68.4
YOLOv10n	2,708,600	8.4	5.50	63.9
YOLO11n	2,590,620	6.4	5.23	67.1
YOLO12	2,550,396	6.5	5.25	63.6
ModelA	2,092,568	5.3	4.40	67.1
ModelB	3,511,190	6.0	7.20	64.3
ModelC	1,948,632	4.4	3.97	63.1
YOLO-Sh	1,710,972	4.1	3.60	68.7
CD-YOLO	2,056,252	5.0	4.24	71.4

According to the comparative experimental results, CD-YOLO achieves highest mAP of 71.4%, significantly outperforming all other compared models. In addition, the parameter count, computational cost, and model size of CD-YOLO are lower than those of most lightweight models. Although ModelC and YOLO-Sh exhibit lower parameter counts, FLOPs, and model sizes than CD-YOLO, their detection performance is substantially inferior. This demonstrates that the proposed CD-YOLO model not only achieves a high degree of lightweight efficiency and is well suited for deployment in resource-constrained environments, but also delivers superior detection performance on the proposed dataset compared with RT-DETR, mainstream YOLO models, and various popular lightweight networks, thereby validating its excellent effectiveness in underwater diseased coral target detection tasks.

3.2.4. Cross-Dataset Adaptability Evaluation

To provide an additional evaluation under different underwater coral datasets, the BHD Coral dataset was introduced to assess the cross-dataset adaptability of the proposed CD-YOLO model. The dataset was first expanded through data augmentation, and then the performance of CD-YOLO and YOLO11n was compared using four evaluation metrics: parameter count, computational cost (FLOPs), model size, and mAP. All experiments were conducted using the same training methodology. The results of the generalization verification experiments are presented in Table 5.

Table 5. Generalization Ability Validation Experiments.

Model	Params	FLOPs/G	Size/MB	mAP/%
YOLO11n	2,590,425	6.4	5.23	82.1
CD-YOLO	2,056,057	5.0	4.24	83.3

The experimental results show that on the BHD Coral dataset, after data augmentation, CD-YOLO still exhibits lower parameter count, computational cost, and model size compared with YOLO11n, while achieving an increase of 1.2 percentage points in mAP. These results indicate that CD-YOLO maintains its lightweight characteristics while exhibiting a certain degree of cross-dataset adaptability under similar underwater coral detection settings.

It should be noted that the experiments on the BHD Coral dataset were conducted under a similar training protocol and with data augmentation applied. Therefore, the results should not be interpreted as a strict cross-dataset generalization test, but rather as preliminary evidence of adaptability under comparable underwater coral detection settings.

3.2.5. Visualization Experiments

To intuitively demonstrate the effect of the model improvements, a visualization experiment was designed. Keeping the experimental environment and training parameters unchanged, the original YOLO11n model and the improved CD-YOLO model were used to detect the same images and generate visual heatmaps. The activation maps were generated using the XGrad-CAM method, which highlights the spatial regions that contribute most to the model's predictions. In the heatmaps, regions colored from blue to red represent areas of model attention, with redder regions indicating higher focus, as shown in Figure 5.

From the experimental results, it can be observed that the original model's attention is dispersed across the entire image. For image (a), the attention is not concentrated on the bleached coral region, with a confidence score of 0.65. For image (d), the attention regions are mainly focused on surface reflections, the seabed, and some non-diseased areas, rather than being concentrated on the bleached coral regions. As a result, bleached corals are misclassified as healthy corals, indicating that the baseline model is prone to distraction by background environments and various interference factors during detection, which leads to misclassification.

In contrast, the improved model demonstrates significantly more focused attention. For image (a), the red regions are concentrated on the bleached coral area, with a higher confidence score of 0.72. Similarly, for image (d), the model accurately focuses on the bleached coral region without misclassification. This behavior suggests that the introduced attention mechanisms enhance the model's ability to suppress irrelevant background features and emphasize disease-related regions. These results indicate that the improvements proposed in this study enable the model to be more sensitive to pathological regions such as bleached coral while reducing the influence of background interference, effectively decreasing the likelihood of false positives and false negatives.

To further intuitively illustrate the detection performance of the proposed model, Figure 6 presents the detection results of the model on multiple images, covering all four target categories. The images include underwater coral scenes with diverse conditions, such as varying illumination, different levels of image clarity, mixed target categories, densely distributed targets, interference from fish, and diseased regions at multiple scales. The results demonstrate that the proposed model can accurately detect all targets across these complex scenarios.

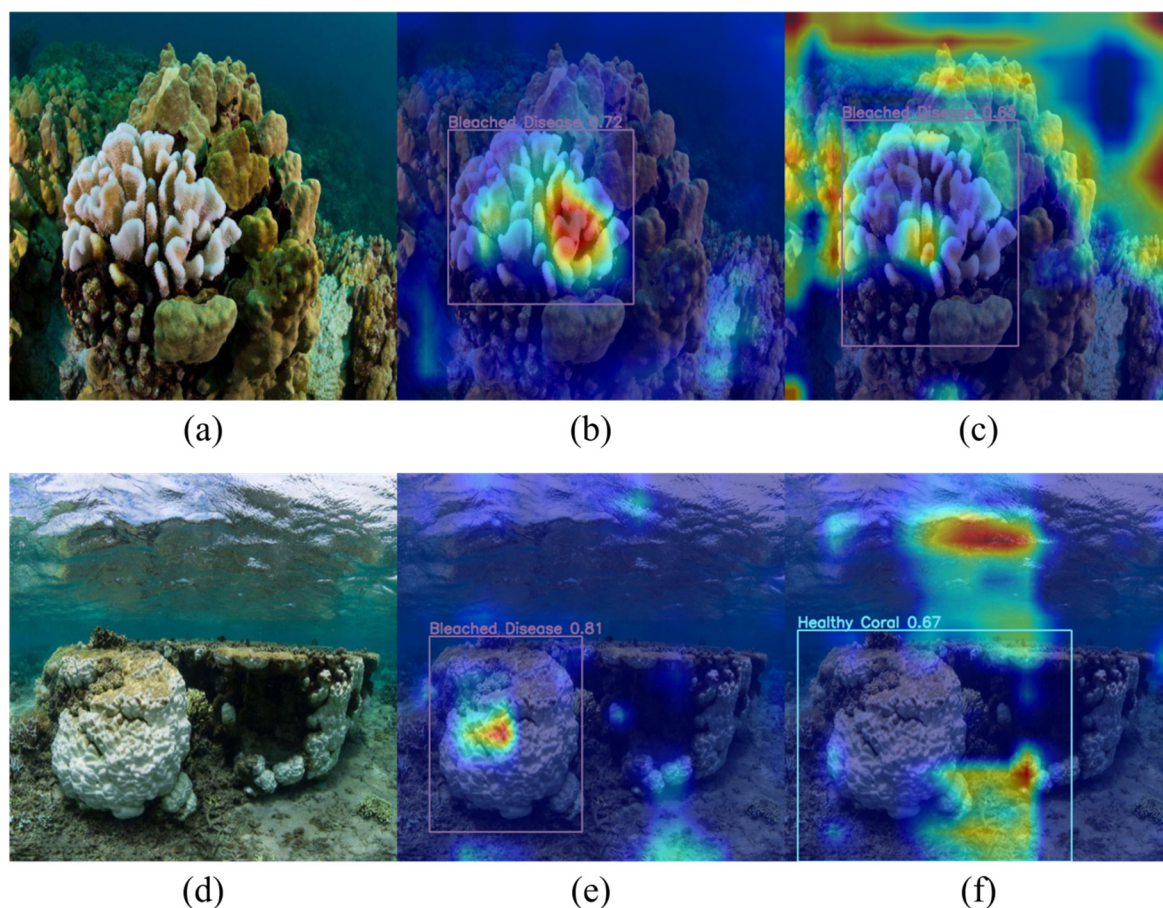


Figure 5. Heat Maps of Focus Areas for the Model Before and After Improvement. (a,d) show the original images from the test set; (b,e) show the visual heatmaps generated by CD-YOLO for (a,d), respectively; (c,f) show the heatmaps generated by YOLO11n for the same images. Among them, (f) illustrates a misclassification case of the baseline model caused by interference from surface reflections, sandy seabed backgrounds, and non-diseased regions.

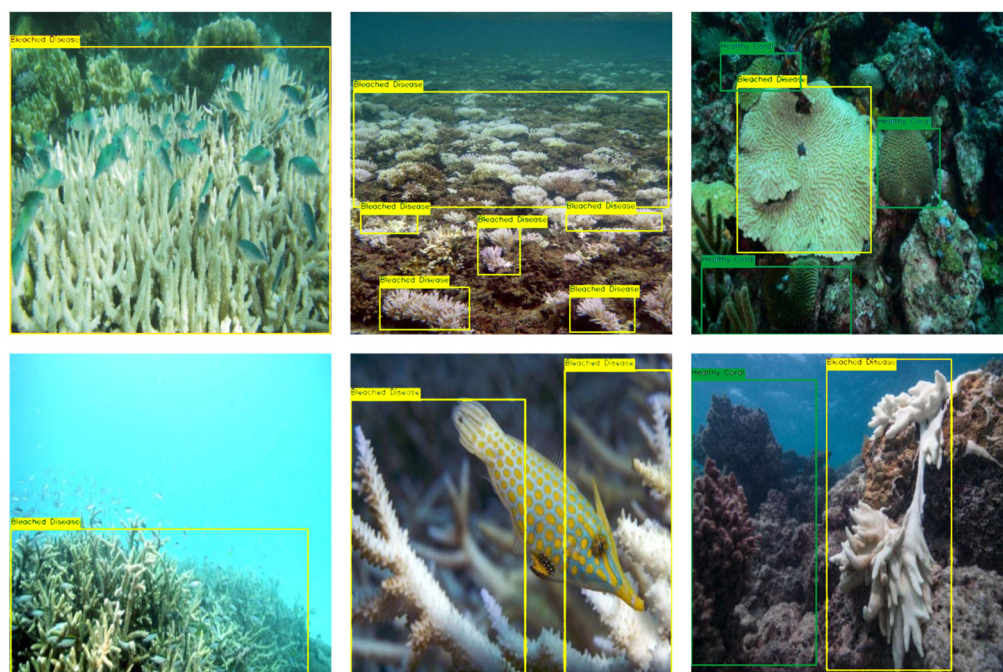


Figure 6. Cont.



Figure 6. Detection Results of the Improved Model on Multiple Images. The bounding boxes indicate the detection results of the model: yellow boxes denote bleached coral, green boxes denote healthy coral, blue boxes denote band disease, and red boxes denote white pox disease.

4. Conclusions

This study addresses the challenges of underwater coral target detection posed by hardware constraints and complex environments, which often lead to high rates of false positives and false negatives. To overcome these challenges, a lightweight yet high-performance underwater coral detection method, CD-YOLO, was proposed. Compared with YOLO11n, CD-YOLO achieves a 4.3 percentage point increase in mAP, while reducing parameter count, computational cost, and model size by 20.6%, 21.9%, and 18.9%, respectively. In comparison with several other lightweight models and related studies, the experimental results indicate that CD-YOLO achieves improved detection performance while maintaining lightweight characteristics. The cross-dataset experiments on the BHD Coral dataset further suggest preliminary adaptability under similar underwater detection conditions. Visualization experiments using heatmaps further illustrate the effectiveness of the proposed improvements.

Due to the lack of access to deployed underwater hardware platforms, real-device latency and energy consumption evaluations were not conducted in this study. Such evaluations will be considered an important part of future work when appropriate hardware conditions become available. Moreover, the conclusions of this study are subject to the limitation of dataset scale, and further validation on larger and more diverse real-world underwater datasets is required.

For future work, collecting a larger dataset encompassing more complex underwater scenarios would allow the model training to better reflect real-world conditions. Additionally, the Neck layer of the model still offers potential for improvement, such as adopting more efficient feature fusion networks or replacing it with advanced upsampling modules to further enhance detection performance.

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

Funding: This research was funded by Shanghai Science and Technology Innovation Program (20dz1203800).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Knowlton, N. The future of coral reefs. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5419–5425. [CrossRef] [PubMed]
2. Moberg, F.; Folke, C. Ecological goods and services of coral reef ecosystems. *Ecol. Econ.* **1999**, *29*, 215–233. [CrossRef]
3. Sutherland, K.P.; Porter, J.W.; Torres, C. Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. *Mar. Ecol. Prog. Ser.* **2004**, *266*, 273–302. [CrossRef]
4. Pandolfi, J.M.; Bradbury, R.H.; Sala, E.; Hughes, T.P.; Bjorndal, K.A.; Cooke, R.G.; McArdle, D.; McClenachan, L.; Newman, M.J.H.; Paredes, G.; et al. Global trajectories of the long-term decline of coral reef ecosystems. *Science* **2003**, *301*, 955–958. [CrossRef]
5. Han, M.W. Environmental Behavior, Sources and Ecological Risks of Polycyclic Aromatic Hydrocarbons in South China Sea Coral Reef Ecosystems. Ph.D. Thesis, Guangxi University, Nanning, China, 2024.
6. Wu, Y.B. Study on the Activity Characteristics of Galactose Lectins from Acropora Corals in Recognizing Zooxanthellae and Pathogens. Master's Thesis, Hainan University, Haikou, China, 2019.
7. Hein, M.Y.; Willis, B.L.; Beeden, R.; Birtles, A. The need for broader ecological and socioeconomic tools to evaluate the effectiveness of coral restoration programs. *Restor. Ecol.* **2017**, *25*, 873–883. [CrossRef]
8. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 386–397. [CrossRef]
9. 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.* **2016**, *39*, 1137–1149. [CrossRef]
10. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A review of YOLO algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
11. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
12. Xin, G.; Xie, H.; Kang, S.; Chen, Y.; Jiang, Y. Improved research on coral bleaching detection model based on FCOS model. *Mar. Environ. Res.* **2024**, *200*, 106644. [CrossRef]
13. González-Rivero, M.; Beijbom, O.; Rodriguez-Ramirez, A.; Bryant, D.E.; Ganase, A.; Gonzalez-Marrero, Y.; Herrera-Reveles, A.; Kennedy, E.V.; Kim, C.J.; Lopez-Marcano, S.; et al. Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. *Remote Sens.* **2020**, *12*, 489. [CrossRef]
14. Wang, L.; Wei, H.; Che, Y.; Zhang, C. Research on coral identification method based on category-adaptive deep data augmentation and transfer learning. *Acta Oceanol. Sin.* **2024**, *46*, 120–130.
15. Hua, M.Z. Research on Coral, Reef Fish and Starfish Recognition Methods Based on Deep Learning. Master's Thesis, Northeast Normal University, Changchun, China, 2023.
16. Zhao, C.; Chen, M. Lightweight underwater benthic organism detection using YOLOv7-RFCW. *Trans. Chin. Soc. Agric. Eng.* **2024**, *40*, 168–177.
17. Li, P.; Li, F.; Ge, Z.; Zhang, T. Underwater object detection algorithm based on improved YOLOv8n. *Electron. Meas. Technol.* **2025**, *48*, 172–179.
18. Ultralytics YOLO11 [EB/OL]. Ultralytics. Available online: <https://docs.ultralytics.com/zh/models/yolo11/#citations-and-acknowledgements> (accessed on 26 February 2025).
19. Xiong, G.; Chen, C.; Zhang, S. QMDF-YOLO11: A Rice Pest Detection Algorithm in Complex Scenarios. *Comput. Eng. Appl.* **2025**, *61*, 113–123.
20. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In *Computer Vision—ECCV 2018*; Springer: Cham, Switzerland, 2018; pp. 112–118.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
22. Selvan, P.S.; Addula, S.R.; Singh, C.E.; Narayanaperumal, M.; Marriwala, N.K.; Appathurai, A. Deep Learning-Enabled Fetal Health Classification Through Sensor-Fused IoT Environment. In *Mobile Radio Communications and 5G Networks*; Marriwala, N.K., Shukla, V.K., Jain, S., Kumar, D., Dhingra, S., Eds.; Lecture Notes in Networks and Systems; Springer: Singapore, 2025; Volume 1328, pp. 157–169.
23. Sunkara, R.; Luo, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Grenoble, France, 19–23 September 2022.
24. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing, Rhodes, Greece, 4–10 June 2023; pp. 1–5.
25. Narayanan, M. SENetV2: Aggregated dense layer for channelwise and global representations. *arXiv* **2023**, arXiv:2311.10807. [CrossRef]
26. Jamil, S.; Rahman, M.; Haider, A. Bag of features (BoF) based deep learning framework for bleached corals detection. *Big Data Cogn. Comput.* **2021**, *5*, 53. [CrossRef]

27. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs beat YOLOs on real-time object detection. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974.
28. Sun, S.; He, L.; Zheng, S.; Xu, X.; Chen, R. Lightweight improvement of YOLOv8n for trash detection under complex background. *J. Electron. Meas. Instrum.* **2025**, *39*, 136–146.
29. Wu, L.; Xu, X. Lightweight tomato leaf pest detection based on improved YOLOv10n. *Smart Agric.* **2025**, *7*, 146–155.
30. Tian, Y.J.; Ye, Q.X.; Doermann, D. YOLOv12: Attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524.
31. Tan, M.X.; Le, Q. EfficientNetV2: Smaller models and faster training. In Proceedings of the 38th International Conference on Machine Learning (ICML 2021), Virtual, 18–24 July 2021; Volume 139, pp. 10096–10106.
32. Zhou, H.; Li, W.; Wei, S.; Men, G.; Wang, Y.; Li, J. Steel surface defect detection method based on YOLOv11-MobileNetv4. *Int. Core J. Eng.* **2025**, *11*, 10–16.
33. Ma, X. Dangerous driving behavior detection algorithm based on Faster-YOLO11. *Inf. Technol. Inf. Educ.* **2025**, 175–178.

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.