*applied sciences*

MDPI

*Article*

# SPyramidLightNet: A Lightweight Shared Pyramid Network for Efficient Underwater Debris Detection

**Yi Luo and Osama Eljamal ***

Water and Environmental Engineering Laboratory, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, 6-1 Kasuga-Koen, Kasuga 816-8580, Fukuoka, Japan
* Correspondence: osama-eljamal@kyudai.jp

**Abstract**

Underwater debris detection plays a crucial role in marine environmental protection. However, existing object detection algorithms generally suffer from excessive model complexity and insufficient detection accuracy, making it difficult to meet the real-time detection requirements in resource-constrained underwater environments. To address this challenge, this paper proposes a novel lightweight object detection network named the Shared Pyramid Lightweight Network (SPyramidLightNet). The network adopts an improved architecture based on YOLOv11 and achieves an optimal balance between detection performance and computational efficiency by integrating three core innovative modules. First, the Split–Merge Attention Block (SMAB) employs a dynamic kernel selection mechanism and split–merge strategy, significantly enhancing feature representation capability through adaptive multi-scale feature fusion. Second, the C3 GroupNorm Detection Head (C3GNHead) introduces a shared convolution mechanism and GroupNorm normalization strategy, substantially reducing the computational complexity of the detection head while maintaining detection accuracy. Finally, the Shared Pyramid Convolution (SPyramidConv) replaces traditional pooling operations with a parameter-sharing multi-dilation-rate convolution architecture, achieving more refined and efficient multi-scale feature aggregation. Extensive experiments on underwater debris datasets demonstrate that SPyramidLightNet achieves 0.416 on the mAP@0.5:0.95 metric, significantly outperforming mainstream algorithms including Faster-RCNN, SSD, RT-DETR, and the YOLO series. Meanwhile, compared to the baseline YOLOv11, the proposed algorithm achieves an 11.8% parameter compression and a 17.5% computational complexity reduction, with an inference speed reaching 384 FPS, meeting the stringent requirements for real-time detection. Ablation experiments and visualization analyses further validate the effectiveness and synergistic effects of each core module. This research provides important theoretical guidance for the design of lightweight object detection algorithms and lays a solid foundation for the development of automated underwater debris recognition and removal technologies.

**Keywords:** underwater object detection; lightweight network; marine debris; deep learning; attention mechanism; feature fusion; real-time detection

check for updates

Academic Editor: Richard Kouzes

Received: 30 July 2025
Revised: 23 August 2025
Accepted: 26 August 2025
Published: 27 August 2025

**Citation:** Luo, Y.; Eljamal, O. SPyramidLightNet: A Lightweight Shared Pyramid Network for Efficient Underwater Debris Detection. *Appl. Sci.* **2025**, *15*, 9404. https://doi.org/10.3390/app15179404

**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

Marine pollution has emerged as one of the most severe global environmental challenges of our time, with underwater debris pollution causing irreversible damage to marine ecosystems and seriously threatening marine biodiversity and the integrity of marine food chains [1]. Artificial debris such as masks, beverage cans, plastic bags, and electronic

devices accumulate extensively on the seafloor, not only destroying marine habitats but also entering the food chain in the form of microplastics, posing potential threats to human health [2,3]. Therefore, developing efficient and accurate underwater debris detection technology holds significant practical importance and urgency for marine environmental protection, ocean cleanup operations, and marine ecological restoration [4].

As a cutting-edge technology at the intersection of computer vision and marine engineering, underwater object detection has attracted widespread attention from both academia and industry in recent years [5,6]. However, the complexity and unique characteristics of underwater environments present unprecedented technical challenges for object detection [7]. First, underwater optical imaging faces physical limitations such as light attenuation, scattering effects, and color distortion, resulting in significantly degraded image quality and blurred target features [8,9]. Second, underwater debris exhibits diverse types, varied morphologies, and dramatic scale variations, often intermixed with complex backgrounds including seafloor sediments and marine organisms, increasing the recognition difficulty for detection algorithms [10]. Finally, underwater operational equipment typically has strict power consumption limitations and computational resource constraints, requiring detection algorithms to possess efficient real-time processing capabilities [11,12]. These challenges render traditional object detection methods ineffective in underwater environments, necessitating the development of targeted technical solutions [13].

The rapid advancement of deep learning technology has provided new technical pathways for underwater object detection [6]. Object detection algorithms based on convolutional neural networks, particularly the You Only Look Once (YOLO) series, have achieved remarkable success in computer vision due to their end-to-end detection framework and excellent real-time performance [14]. YOLOv11 has demonstrated significant improvements in network architecture design, feature extraction capabilities, and detection accuracy [15]. However, existing YOLOv11 algorithms still face critical issues in underwater debris detection applications. On the one hand, excessive model complexity leads to computational resource consumption that limits deployment on resource-constrained underwater equipment [16]. On the other hand, standard feature extraction and fusion modules struggle to effectively address the special imaging conditions of underwater environments, affecting detection accuracy and robustness [17]. Achieving model lightweighting while maintaining detection accuracy has become a core challenge urgently requiring resolution in the underwater object detection field [18].

Existing lightweight methods primarily include strategies such as network pruning, knowledge distillation, model quantization, and efficient network architecture design [19]. While these methods can reduce model parameters and computational complexity to some extent, they are often accompanied by significant drops in detection accuracy, particularly in complex underwater environments where simple model compression strategies may lose critical feature information [20]. Moreover, although traditional feature pyramid structures can handle multi-scale targets, their pooling-based design tends to cause detail information loss, while the parameter redundancy of spatial pyramid pooling modules also constrains model efficiency optimization [21,22]. Therefore, designing innovative architectures that can significantly reduce model complexity while maintaining or even improving detection accuracy has become a key research challenge [23].

To address these technical challenges, this paper proposes a novel lightweight underwater debris detection network, the Shared Pyramid Lightweight Network (SPyramidLightNet). The core contribution of this network lies in the collaborative design of three innovative modules: the Split–Merge Attention Block (SMAB) achieves efficient multi-scale feature extraction through dynamic mixed convolution and split–merge attention mechanisms; the C3 GroupNorm Detection Head (C3GNHead) significantly reduces detection head

computational overhead through parameter sharing and GroupNorm normalization strategies; and the Shared Pyramid Convolution (SPyramidConv) realizes refined multi-scale feature aggregation through a weight-sharing multi-dilation-rate convolution architecture. This innovative modular combination strategy not only resolves the dilemma between lightweighting and accuracy in traditional methods but also incorporates optimized design specifically tailored for the unique characteristics of underwater environments.

The main contributions of this paper can be summarized as follows:

(1) The SMAB module is proposed, enhancing feature representation while reducing parameters through dynamic weight allocation and multi-scale feature fusion.
(2) The C3GNHead lightweight detection head is designed, significantly reducing computational complexity via shared convolution and GroupNorm optimization.
(3) The SPyramidConv module is developed, achieving efficient multi-scale feature extraction through parameter-sharing dilated convolution architecture.
(4) The complete SPyramidLightNet architecture is constructed, achieving optimal balance between lightweighting and high-precision detection through synergistic module optimization. Extensive experiments on underwater debris datasets validate the method's effectiveness and superiority, providing technical support for underwater environmental protection and ocean cleanup technologies.

## 2. Related Work

The remarkable advancement of deep learning technologies has revolutionized computer vision applications, leading to substantial improvements in object detection accuracy [24]. However, this progress has been accompanied by a significant escalation in model complexity and computational demands, which poses considerable challenges for deployment in resource-constrained environments [25]. This limitation becomes particularly pronounced in specialized application domains such as underwater object detection, where computational resources, memory capacity, and power consumption are subject to stringent constraints, thereby necessitating the development of lightweight alternatives [26]. This section presents a comprehensive review of the state-of-the-art research from three complementary perspectives: lightweight neural network architectures, efficient object detection algorithms, and underwater object detection methodologies [27]. This systematic analysis establishes the theoretical foundation that underpins the technical innovations presented in this paper [28].

### *2.1. Lightweight Neural Network Design*

The emergence of lightweight neural networks stems from the practical demands of mobile devices and edge computing applications, with the primary objective of substantially reducing computational complexity and storage requirements while maintaining model performance through architectural optimization and parameter compression techniques [26,29]. Initial efforts in this domain predominantly focused on post-processing approaches such as network pruning and knowledge distillation [20]; however, recent research has shifted towards developing inherently efficient architectures at the design stage [30,31].

The MobileNet family represents a pivotal advancement in lightweight network design. MobileNetV1 [29] pioneered the adoption of depthwise separable convolutions, which decompose standard convolutions into depthwise and pointwise operations, achieving dramatic reductions in both parameter count and computational cost while preserving competitive performance. Building upon this foundation, MobileNetV2 [30] introduced the inverted residual block architecture with linear bottlenecks, optimizing feature propagation efficiency through a strategic combination of low-dimensional compression and high-

dimensional expansion. MobileNetV3 [32] further advanced the field by incorporating Neural Architecture Search (NAS) techniques alongside Squeeze-and-Excitation attention mechanisms, enabling more refined architectural optimization.

The ShuffleNet series explores lightweight design from the perspective of channel shuffling operations. ShuffleNetV1 [33] leverages group convolutions combined with channel shuffle operations to maintain representational capacity while significantly reducing computational complexity. ShuffleNetV2 [34] subsequently reevaluated lightweight design principles from the standpoint of actual inference speed, establishing design guidelines that minimize Memory Access Cost (MAC) and highlighting the discrepancy between theoretical computation and practical inference performance.

The EfficientNet [35] series introduced a systematic approach through compound scaling strategies, which balance the scaling relationships across network depth, width, and resolution dimensions. This work pioneered the theoretical analysis of optimal scaling ratios across different network dimensions, providing crucial design principles for lightweight architectures. The subsequent EfficientNetV2 [36] further incorporated progressive learning strategies and refined architectural improvements, enhancing training efficiency while maintaining computational efficiency.

### *2.2. Lightweight Object Detection Algorithms*

The complexity of object detection tasks poses significant challenges for lightweight design, requiring coordinated optimization across multiple components including backbone networks, feature fusion modules, and detection heads [37,38]. Current lightweight object detection methods can be categorized into three main technical approaches: network architecture optimization, feature compression, and computational acceleration [23,27].

In network architecture optimization, the YOLO series [14,39–41] has played a pivotal role in lightweight object detection. YOLOv3 achieves an effective balance between accuracy and speed through its Darknet-53 backbone and Feature Pyramid Network (FPN) structure. YOLOv4 enhances detection performance by incorporating Cross Stage Partial (CSP) structures and PANet feature fusion strategies. The YOLOv5 series offers various model scales through modular design and hyperparameter optimization, addressing diverse deployment requirements. Recent YOLO variants, including YOLOv6, YOLOv7, and YOLOv8, explore lightweight optimization strategies from different perspectives, such as re-parameterization techniques, gradient flow optimization, and architecture search.

The Single Shot MultiBox Detector (SSD) [42] implements efficient single-stage detection through multi-scale feature maps and prior box mechanisms. The MobileDet series [43] combines MobileNet with SSD, specifically optimized for mobile devices. RetinaNet [44] addresses class imbalance in single-stage detection through Focal Loss, spawning various lightweight variants.

Regarding feature compression techniques, FPN [45] and its variants such as PANet [46] and BiFPN [47] provide efficient solutions for multi-scale feature fusion. These methods reduce computational overhead while maintaining detection accuracy by optimizing feature propagation paths and fusion strategies. Recent studies like Weighted Bi-directional FPN [48] further enhance fusion efficiency through learnable weight allocation mechanisms.

Computational acceleration techniques primarily encompass quantization, pruning, and knowledge distillation. Quantization significantly reduces model storage and computational requirements by lowering numerical precision (e.g., INT8 quantization). Network pruning achieves model compression by removing redundant weights or neurons. Knowledge distillation transfers knowledge from large networks to compact ones through teacher-student training strategies.

### 2.3. Underwater Object Detection Algorithms

Image enhancement methods attempt to preprocess underwater images before detection, improving image quality and enhancing subsequent detection algorithm performance. Traditional enhancement methods include histogram equalization, homomorphic filtering, Retinex theory, and dark channel prior [49]. While computationally simple, these methods typically address only specific image degradation issues and lack generalizability [50].

Recent years have witnessed significant advances in deep learning-based image enhancement methods. Wu et al. [51] proposed FW-GAN, which learns the mapping from underwater images to clear images through generative adversarial networks, simultaneously addressing color distortion and contrast degradation. Yin et al. [52] incorporated physical constraints, guiding network learning for more accurate image restoration by modeling underwater imaging physics. Li et al. [53] developed FUnIE-GAN, specifically optimized for underwater image color correction and clarity enhancement, achieving excellent results across multiple underwater datasets.

Network architecture adaptation methods modify existing object detection network structures or components to better accommodate underwater environmental characteristics. The core principle involves localized optimization for underwater image properties while maintaining the basic detection framework structure. The introduction of attention mechanisms represents a crucial development direction. Chen et al. [54] proposed Water-Net, embedding channel and spatial attention modules into Faster R-CNN's backbone, enhancing the network's focus on critical underwater target features. This method adaptively adjusts feature channel weights, effectively suppressing noise interference in underwater environments. Li et al. [55] designed Underwater-YOLO, incorporating multi-scale attention mechanisms into YOLOv3, improving small object detection through dynamic weight allocation across different scale feature maps.

Feature fusion strategy improvement constitutes another important research direction. Chen et al. [56] optimized feature fusion paths to enhance multi-scale feature connections, introduced multi-scale adaptive fusion strategies for improved detection accuracy, and reparameterized feature extraction networks for computational simplification. Yang et al. [57] proposed LAFFNet, employing an encoder–decoder structure with multiple AAF modules as its core. Each AAF module contains multi-branch convolutions for multi-scale feature extraction and achieves adaptive feature fusion through channel attention mechanisms. Gong et al. [58] introduced a color feature fusion-based underwater image enhancement method, performing multi-scale convolutional feature extraction on RGB channels separately to effectively learn global and local image information. They designed residual enhancement modules with attention mechanisms to improve feature representation capability, while constructing dynamic feature enhancement modules through deformable convolutions to enhance modeling accuracy for complex underwater scenes.

Comprehensive analysis reveals that existing underwater object detection methods primarily focus on improving detection accuracy, with insufficient attention to lightweight design. This status quo contradicts the practical requirements of underwater application scenarios: underwater equipment typically faces strict constraints on computational capacity, storage space, and power consumption, necessitating efficient and lightweight algorithms. Therefore, achieving significant model lightweighting while maintaining detection accuracy has become a critical challenge in underwater object detection.

## 3. Methodology

To address the dual challenges of excessive model complexity and insufficient detection accuracy in underwater object detection, this paper presents a novel lightweight detection network, the Shared Pyramid Lightweight Network (SPyramidLightNet). By inte-

grating three core innovative modules, this network significantly reduces model parameters and computational complexity while effectively enhancing object detection performance in complex underwater environments. As illustrated in Figure 1, SPyramidLightNet adopts an improved architecture based on YOLOv11, constructing an efficient and accurate underwater object detection framework through key technical innovations including the Split-Merge Attention Block (SMAB), C3 GroupNorm Detection Head (C3GNHead), and Shared Pyramid Convolution (SPyramidConv).

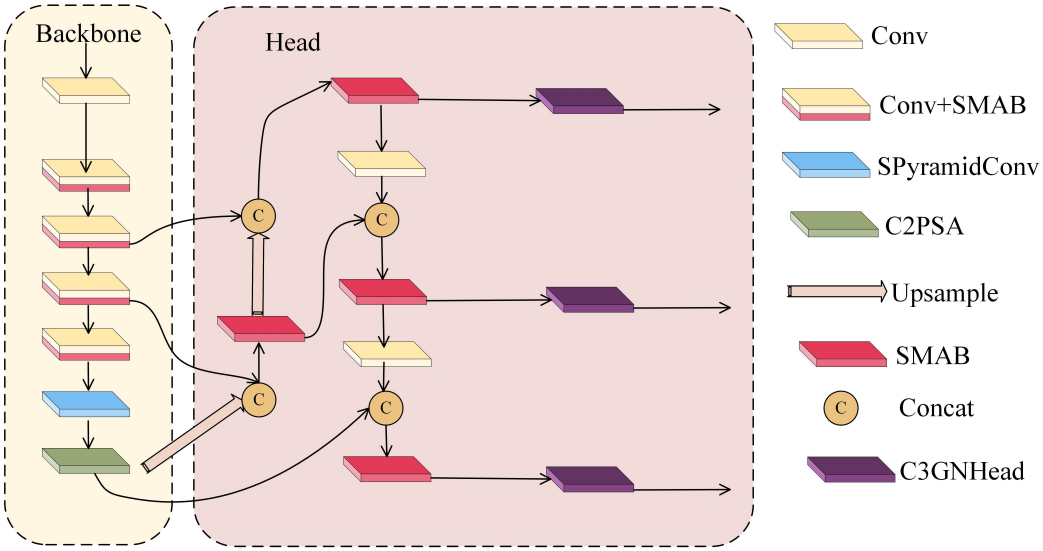



**Figure 1.** Overview of the SPyramidLightNet Model Architecture.

SPyramidLightNet follows the classical encoder–decoder design paradigm, comprising two core components: a backbone network and detection head. The backbone employs a progressive feature extraction strategy, achieving efficient multi-scale feature extraction through multiple convolutional layers and the innovatively designed SMAB module. Specifically, the SMAB module enhances feature representation capability while significantly reducing parameters through dynamic mixed convolution and split–merge attention mechanisms. The SPyramidConv module implements refined and efficient multi-scale feature aggregation via a parameter-sharing multi-dilation-rate convolutional architecture. The C2PSA module further strengthens the network's global context modeling capability.

In the detection head, the network adopts a Feature Pyramid Network (FPN) structure, effectively integrating multi-scale features through upsampling and feature fusion operations. The innovative C3GNHead ultimately accomplishes lightweight object classification and localization tasks. This detection head maintains detection accuracy while reducing computational overhead through GroupNorm normalization and shared convolution mechanisms. Overall, SPyramidLightNet successfully achieves an optimal balance between detection performance and computational efficiency through these core technical innovations, providing an excellent solution for real-time object detection in resource-constrained environments.

### *3.1. Split–Merge Attention Block (SMAB)*

While the C3K2 module widely used in traditional YOLOv11 architectures demonstrates good feature extraction performance, its fixed convolutional structure and extensive parameter configuration face challenges of low computational efficiency and limited feature representation capability when processing complex underwater scenes. Particularly in underwater debris detection tasks, where target scales vary dramatically, morphologies are diverse, and backgrounds are complex, traditional C3K2 modules struggle to achieve

model lightweighting while maintaining detection accuracy. To address this challenge, we propose an innovative Split–Merge Attention Block (SMAB) module that significantly reduces parameters while enhancing the model's adaptability and detection accuracy in complex underwater environments through dynamic kernel selection mechanisms and multi-scale feature adaptive fusion strategies.

As illustrated in Figure 2, the core design philosophy of the SMAB module centers on efficient feature learning through a split-process-merge strategy, primarily comprising two key components: Enhanced Inception Depthwise Convolution (EIDWConv) and EIDWConv Mixer. The module's forward propagation process can be described as a multi-stage feature transformation sequence: first splitting the input feature map along the channel dimension, then extracting features through parallel multi-scale convolutional branches, and finally completing information integration through attention-guided feature fusion.

As the core component of SMAB, EIDWConv achieves adaptive selection of different scale convolution kernels through a dynamic weight allocation mechanism. Given an input feature map $X_{C\times H\times W}$, EIDWConv first generates dynamic weights through global average pooling and $1 \times 1$ convolution:

$$W_{akw} = Softmax(Reshape(\sigma(Conv_{1\times1}(GAP(X)))), [3, B, C, 1, 1]) \tag{1}$$

where $GAP$ denotes the global average pooling operation, $\sigma$ represents the activation function, and the dynamic weights $W_{dkw}in[3, B, C, 1, 1]$ are used to modulate the contributions of three different convolution kernels. Subsequently, three parallel depthwise convolution branches process the input features, respectively:

$$F_{total} = \sum_{i=0}^{2} W_{akw}^{(i)} \bigodot DWConv^{(i)}(X) \tag{2}$$

In this context, $DWConv^{(i)}$ represents depthwise convolution operations employing $i \times i$ square kernels, $1 \times (3k + 2)$ horizontal strip kernels, and $(3k + 2) \times 1$ vertical strip kernels, with $\odot$ denoting element-wise multiplication.

$EIDWConv$ Mixer extends the capability of multi-scale feature extraction through a channel-splitting strategy that enables more efficient feature learning. For input features X, the method first performs equal partitioning along the channel dimension:

$$X_1, X_2 = Split(X, dim = 1) \tag{3}$$

Each group is processed through $EIDWConv$ with different configurations, followed by feature fusion via $1 \times 1$ convolution:

$$Y_{\text{mixer}} = Conv_{1\times1}(Concat(EIDWConv_1(X_1), EIDWConv_2(X_2))) \tag{4}$$

This grouping strategy not only reduces computational complexity but also enhances feature representation diversity through the combination of convolution kernels at different scales. Here, $EIDWConv_1(X_1), EIDWConv_2(X_2)$ employs various kernel configuration parameters, enabling the network to simultaneously capture both fine-grained and coarse-grained feature information. The cross-channel information exchange and feature integration are ultimately achieved through $1 \times 1$ convolutions.

The SMAB module successfully addresses the fundamental limitations of traditional C3K2 modules in underwater target detection tasks through its carefully designed multi-scale dynamic convolution architecture. This module not only significantly reduces parameter count and computational complexity but also substantially enhances the model's adaptability to complex underwater environments through dynamic weight allocation

and multi-scale feature fusion mechanisms. Notably, its innovative split–parallel–merge strategy and dynamic kernel selection mechanism enable the network to adaptively adjust feature extraction strategies based on varying input characteristics. This approach maintains high detection accuracy while achieving remarkable model compression, providing an efficient and robust solution for underwater debris detection tasks.

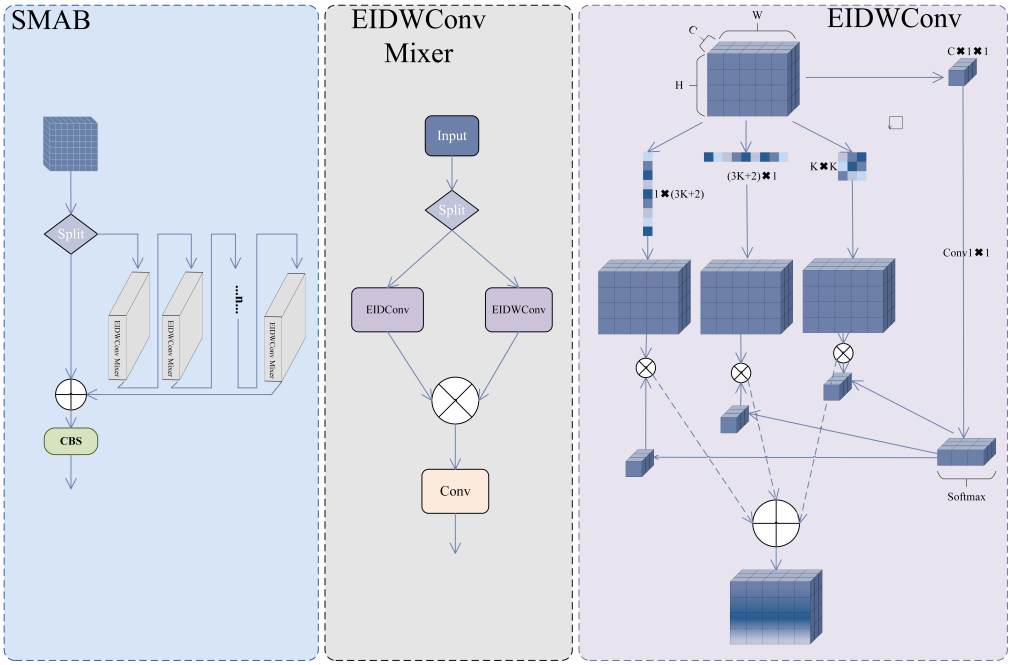



**Figure 2.** Overall Structural Diagram of the SMAB Module.

### 3.2. C3 GroupNorm Detection Head (C3GNHead)

Traditional YOLOv11 detection heads employ independent convolutional layers to process feature maps at different scales, while this design ensures detection accuracy, it suffers from parameter redundancy and excessive computational complexity. In resource-constrained applications such as underwater debris detection, the computational overhead of detection heads often constitutes a substantial portion of the model's total computation, severely limiting real-time performance and deployment efficiency. To address this critical challenge, we propose the C3 GroupNorm Detection Head (C3GNHead), which introduces shared convolution mechanisms, GroupNorm normalization strategies, and scale-adaptive adjustment techniques. This approach significantly reduces parameters and computational complexity while effectively maintaining or even improving detection accuracy. The design not only resolves the efficiency bottleneck of conventional detection heads but also incorporates optimizations tailored to underwater environments, offering an efficient and practical solution for lightweight object detection.

As illustrated in Figure 3, the core design philosophy of C3GNHead is to achieve lightweight detection head improvements through parameter sharing and normalization optimization. The entire detection head processing pipeline can be divided into three key stages: feature alignment, shared feature extraction, and multi-task output. Initially, feature maps $F_{t\in\{1,2,3\}}$ from different scales undergo channel dimension unification through independent $1 \times 1$ convolutions and GroupNorm, ensuring the feasibility of subsequent shared convolutions. For multi-scale input feature maps, the feature alignment process can be formulated as:

$$F_t^{align} = SiLU(GN(Conv_{1\times1}^{(i)}(F_i))), i = 1, 2, 3 \quad (5)$$

In this formulation, $Conv_{1\times1}^{(i)}$ denotes the $1 \times 1$ convolution associated with the $i$-th scale, and $GN$ represents the GroupNorm operation configured with 16 groups to optimize the trade-off between normalization quality and computational cost. GroupNorm demonstrates enhanced stability over conventional BatchNorm in small-batch training and inference scenarios, a critical advantage for detection applications.

After feature alignment, all scale-specific feature maps proceed to the shared C3GN module for comprehensive feature extraction. By combining depthwise and pointwise convolutions, this module preserves feature extraction capabilities while achieving significant model complexity reduction via parameter sharing. The shared feature extraction procedure is formulated as:

$$F_i^{shared} = Conv_{1\times1}(SiLU(GN(DWConv_{3\times3}(F_t^{align})))) \tag{6}$$

In this context, $DWConv_{3\times3}$ denotes $3 \times 3$ depthwise convolution, achieving depthwise separable convolution through configuring the groups parameter to match input channel count, thereby substantially reducing parameters. Following this, $1 \times 1$ convolution performs cross-channel information integration to recover feature expressiveness.

During the multi-task output phase, C3GNHead produces distinct regression and classification predictions. To address target scale differences across detection layers, a learnable Scale layer is incorporated for adaptive adjustment of regression branch outputs. The prediction generation procedure is formulated as:

$$P_{reg}^{(i)} = r_i \bigodot Conv_{reg}(F_i^{shared}) \tag{7}$$

$$P_{cls}^{(i)} = Conv_{cls}(F_i^{shared}) \tag{8}$$

In this formulation, $\gamma_i$ denotes the learnable scale parameter for the $i$-th detection layer with an initial value of 1.0, and $Conv_{reg}$ and $Conv_{cls}$ indicate the $1 \times 1$ convolutional layers designated for regression and classification tasks. By incorporating the Scale layer, the network autonomously learns appropriate output magnitudes for various scales, addressing scale discrepancies that may result from utilizing shared convolutions.

Feature concatenation yields the final detection output:

$$O_i = Concat(p_{reg}^{(i)}, p_{cls}^{(i)}) \in R^{B\times(4\times R+C)\times H_i\times W_i} \tag{9}$$

In this context, $R$ denotes the regression maximum (commonly set to 16), $C$ represents the category count, and $H_i$ and $W_i$ correspond to the height and width dimensions of the $i$-th scale feature map. This architecture preserves multi-scale detection completeness while achieving substantial computational efficiency gains via parameter sharing.

GroupNorm implementation additionally strengthens detection head stability. In contrast to BatchNorm's dependence on batch-wise statistics, GroupNorm partitions channels into groups with intra-group normalization, rendering the process batch-size independent—a crucial advantage in detection applications:

$$GN(X) = \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} \bigodot \gamma + \beta \tag{10}$$

In this formulation, $\mu$ and $\sigma$ denote the intra-group mean and variance, whereas $\gamma$ and $\beta$ correspond to learnable parameters for affine transformation.

Through its carefully engineered lightweight architecture, C3GNHead successfully addresses the fundamental limitations of traditional detection heads in terms of parameter efficiency and computational complexity. The introduction of shared convolution

mechanisms enables parameter reuse across different scales for feature extraction, substantially reducing model storage requirements. The adoption of GroupNorm not only improves normalization stability but also enhances the model's adaptability to varying batch sizes. The innovative Scale layer design effectively compensates for scale sensitivity issues that may arise from parameter sharing. Overall, C3GNHead significantly improves computational efficiency and deployment feasibility while maintaining detection accuracy, providing an efficient and reliable technical solution for real-time object detection applications in resource-constrained scenarios such as underwater debris detection.

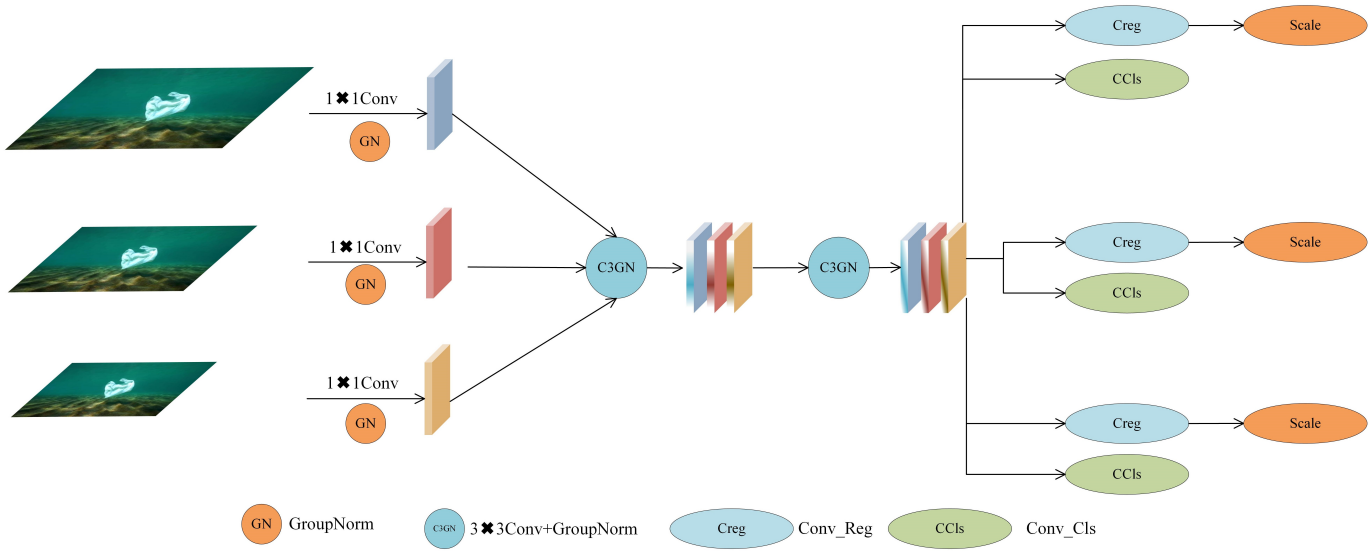



**Figure 3.** Overall Architecture of the C3GNHead Module.

### *3.3. Shared Pyramid Convolution (SPyramidConv)*

The Spatial Pyramid Pooling Fast (SPPF) module widely adopted in traditional YOLOv11 architectures, while capable of effectively aggregating multi-scale information, suffers from information loss and limited feature representation capacity due to its pooling-based design. In complex scenarios such as underwater debris detection, the downsampling nature of pooling operations often discards critical detail information. This loss becomes particularly detrimental given the blurred boundaries and complex textures characteristic of underwater targets. To address this fundamental challenge, we propose the Shared Pyramid Convolution (SPyramidConv) module, which introduces a parameter-sharing multi-dilation-rate convolution architecture. This approach maintains multi-scale feature extraction capabilities while significantly enhancing feature representation precision and computational efficiency. The module not only overcomes the inherent limitations of traditional pooling operations but also achieves parameter optimization through an innovative weight-sharing mechanism, providing a stronger feature representation foundation for underwater object detection.

As illustrated in Figure 4, the core design philosophy of SPyramidConv is to construct a feature pyramid through shared-weight convolutions with multiple dilation rates, achieving efficient multi-scale feature extraction. The entire module's processing pipeline encompasses three key stages: channel reduction, multi-scale feature extraction, and feature fusion. Initially, input feature map $X_{C\times H\times W}$ undergoes channel reduction via $1 \times 1$ convolution to decrease computational complexity and prepare for subsequent multi-branch processing. The channel reduction process can be expressed as:

$$F_0 = Conv_{1\times1}(X) \in R^{B\times C/2\times H\times W} \tag{11}$$

In this formulation, $Conv_{1\times1}$ denotes the $1 \times 1$ convolutional operation that halves the input channel count to reduce downstream computational demands. This dimension reduction approach preserves feature expressiveness while providing an efficient basis for parallel multi-branch operations.

Multi-scale feature extraction constitutes the core innovation of SPyramidConv. Unlike traditional methods employing independent kernels, this module adopts a weight-sharing strategy, achieving feature extraction across different receptive fields by adjusting dilation rates. Given a set of dilation rates $D = \{d_1, d_2, \ldots, d_i\}$, the multi-scale feature extraction process can be recursively defined as:

$$F_i = DilatedConv(F_{i-1}; W_{shared}, d_i), i = 1, 2, \ldots, k \tag{12}$$

In this context, $W_{shared}$ denotes the shared convolution weights utilized by all dilation branches, while $d_i$ indicates the dilation rate for the $i$-th branch. The dilated convolution can be mathematically expressed as:

$$DilatedConv(F; W, d)[m, n] = \sum_{i,j} F[m + d \cdot i, n + d \cdot j] \cdot W[i, j] \tag{13}$$

This recursive feature extraction strategy enables each branch to progressively expand its receptive field based on the previous branch, thereby constructing hierarchical multi-scale feature representations. The incremental design of dilation rates ensures progressive feature capture from local details to global context.

To ensure effective convolution operations across different dilation rates, the padding size for each branch requires adaptive adjustment based on the dilation rate:

$$p_i = \frac{d_i \cdot (k-1) + 1}{2} \tag{14}$$

In this formulation, $k$ denotes the convolution kernel size (commonly 3), while $p_i$ indicates the necessary padding for the $i$-th branch. Through this adaptive padding approach, output feature maps maintain uniform spatial dimensions, enabling seamless feature fusion in later stages.

During the feature fusion phase, all branch outputs and the initial dimensionality-reduced features undergo channel-wise concatenation, with a subsequent $1 \times 1$ convolution performing the final feature consolidation:

$$Y = Conv_{1\times1}(Concat(F_0, F_1, F_2, \ldots, F_k)) \tag{15}$$

This design not only preserves direct propagation paths for original features but also integrates contextual information across different scales, ultimately forming rich and hierarchical feature representations. The introduction of parameter sharing mechanisms enables the model to obtain multi-scale features while avoiding parameter redundancy inherent in traditional multi-branch architectures.

The core advantage of the weight-sharing mechanism lies in its ability to reuse the same feature extractor across different receptive field scales, which not only substantially reduces model parameters but also enhances feature extraction consistency. The shared weights $W_{shared}$ learn universal feature extraction patterns through end-to-end training, enabling the network to maintain unified and effective feature representation across different scales:

$$W^*_{shared} = argmin\, w \sum_{i=1}^{k} \ell(DilatedConv(F_{i-1}; W, d_i), T_i) \tag{16}$$

In this formulation, $\ell$ denotes the loss function, while $T_i$ indicates the target feature representation for the $i$-th scale.

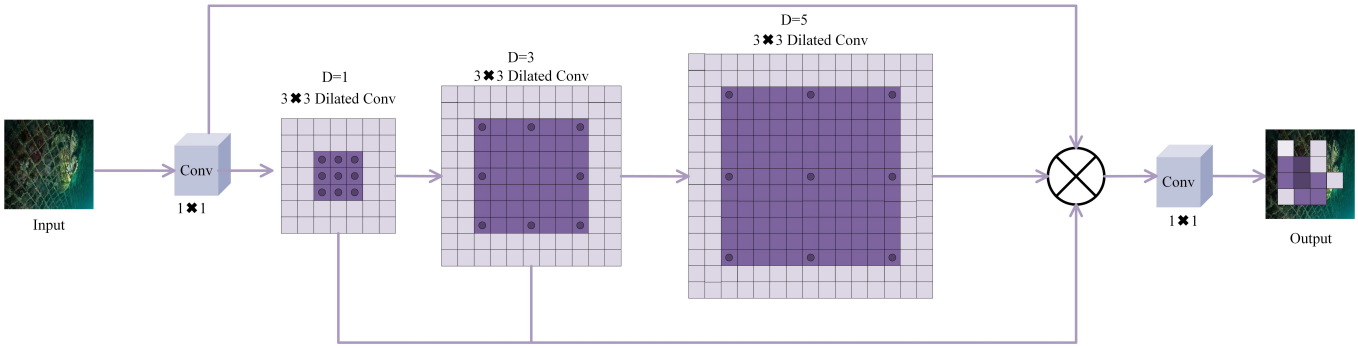



**Figure 4.** Structural Diagram of the SPyramidConv Module.

Through its meticulously engineered shared pyramid convolution architecture, SPyramidConv successfully addresses the fundamental limitations of traditional SPPF modules in terms of feature fidelity and computational efficiency. Unlike the information compression inherent in pooling operations, convolutions better preserve spatial detail information—a critical factor for precise target localization in underwater debris detection. The implementation of parameter sharing not only significantly reduces model complexity but also enhances feature consistency across scales through unified feature extraction patterns. The recursive multi-scale design enables the network to construct richer feature hierarchies, with progressive feature capture from local textures to global semantics providing a robust feature foundation for object detection in complex underwater scenarios. Overall, SPyramidConv achieves substantial efficiency optimization while maintaining or even enhancing feature representation capabilities, offering crucial technical support for the practical deployment of underwater debris detection algorithms.

## 4. Results and Analysis

### *4.1. Dataset Description and Experimental Setup*

#### 4.1.1. Dataset Description

This study employs an underwater debris detection dataset to validate the effectiveness of SPyramidLightNet. The dataset encompasses 15 debris categories: medical masks, metal cans, cellphones, electronics, glass bottles, gloves, metal objects, miscellaneous waste, fishing nets, plastic bags, plastic bottles, plastic products, metal rods, sunglasses, and tires. These categories represent the most prevalent types of marine debris, holding significant environmental protection importance.

Image acquisition covers diverse underwater environmental conditions, including varying water depths, illumination levels, water clarity, and seabed topography. Image resolutions range from 640 × 480 to 1920 × 1080, authentically reflecting typical underwater challenges: low contrast from light attenuation, image blur caused by water scattering, color distortion, and complex background interference. As illustrated in Figure 5, plastic debris dominates the dataset in terms of category characteristics. Metal debris exhibits relatively clear boundaries but is easily confused with seabed rocks. Electronic waste presents complex and varied shapes. Irregularly shaped debris such as fishing nets highly blend with the seabed environment, posing detection challenges.

The labeled images were divided into training, validation, and testing subsets in a ratio of 8:1:1, corresponding to 4000, 500, and 500 images, respectively. Data augmentation, including random rotation, flipping, scaling, and brightness/contrast adjustment, was applied during training. The partitioning process maintains balanced category distribution

across different subsets while considering temporal and spatial information of image acquisition, ensuring appropriate environmental diversity between training and test sets to better evaluate algorithmic generalization capabilities. Additionally, the dataset covers a wide range of water depths, from approximately 3 m to 25 m, including shallow coastal zones and moderately deep marine environments. This variation in depth enhances the environmental diversity of the dataset and contributes to the robustness of the model evaluation under different real-world underwater conditions. In terms of object scale, the dataset includes debris of varying sizes, with the smallest annotated objects occupying approximately 16 × 16 pixels, or around 0.5% of the total image area. SPyramidLightNet is capable of reliably detecting debris targets at this minimal size, demonstrating its effectiveness in identifying small and visually subtle objects in complex underwater environments. This dataset provides a standardized evaluation benchmark for underwater object detection algorithms, enabling comprehensive validation of the proposed algorithm's effectiveness in real marine environments.



**Figure 5.** Example images of targets in an underwater dataset.

#### 4.1.2. Experimental Setup

To objectively evaluate the comprehensive performance of SPyramidLightNet in underwater debris detection tasks, this study establishes a systematic experimental evaluation framework, conducting comprehensive performance validation across multiple critical dimensions including detection accuracy, model complexity, inference speed, and robustness. The experimental design adheres to rigorous scientific principles, encompassing quantitative comparative analysis with current mainstream lightweight object detection algorithms, ablation studies of key modules, and generalization performance evaluation under various underwater environmental conditions, ensuring objectivity, reliability, and scientific value of experimental results.

All experiments in this research were conducted on a standardized high-performance computing platform to ensure consistency of experimental conditions and reproducibility of results. The experimental environment was built on the Windows 11 operating system, with a hardware platform configured as a single-GPU high-performance architecture featuring an NVIDIA GeForce RTX 4060 (NVIDIA, Santa Clara, CA, USA) graphics card equipped with 8GB GDDR6 high-speed memory, providing sufficient computational resources for efficient deep neural network training and inference. The software development environment was constructed based on Python 3.10 programming language, with PyTorch 2.0.1 as the core deep learning framework, utilizing the CUDA 11.8 parallel computing toolkit for GPU acceleration optimization, and integrating the OpenCV 4.8.0 computer vision library for image preprocessing and post-processing operations, ensuring platform stability and development efficiency.

Considering the specificity of underwater debris detection tasks and the training characteristics of lightweight models, this study employs carefully optimized training strategies and hyperparameter configuration schemes. The training process spans 300 epochs, ensuring sufficient learning and convergence in complex underwater scenarios. The batch size is set to 16, maximizing GPU parallel computing capabilities while maintaining gradient estimation stability. The AdamW optimizer is selected for its decoupled weight decay mechanism that better adapts to lightweight network training characteristics, with initial learning rate set to 0.001 and weight decay coefficient configured as 0.0001, effectively preventing overfitting under limited parameter conditions.

The learning rate scheduling strategy adopts a multi-stage adaptive adjustment scheme, combining warm-up initialization, multi-step decay, and minimum learning rate protection mechanisms. During the first 5 epochs, a linear warm-up strategy gradually increases the learning rate from $1 \times 10^{-6}$ to the initial 0.001, ensuring smooth parameter initialization. Subsequently, at epochs 100, 200, and 250, the learning rate is adjusted by a decay factor of 0.1, with the final learning rate floor set at $1 \times 10^{-6}$. This strategy effectively accommodates the training dynamics of lightweight networks, achieving refined parameter optimization while maintaining convergence speed.

### *4.2. Visual Result Analysis*

#### 4.2.1. Visualization and Validation of Feature Maps

To comprehensively understand SPyramidLightNet's feature-learning mechanisms and hierarchical representation capabilities in underwater debris detection tasks, this study conducts systematic visualization analysis of feature maps across different network stages. By selecting key layers from the backbone architecture (Stage 0, 2, 4, 6, 9, 16, 19, 22) and employing activation heatmap visualization, we intuitively demonstrate SPyramidLightNet's progressive learning process from shallow detail features to deep semantic representations. This visualization analysis not only validates the rationality of the network

architecture design but also provides crucial empirical evidence for understanding the functional mechanisms of innovative modules during feature extraction.

As illustrated in Figure 6, visualization results of shallow features reveal that Stage 0 and Stage 2 feature maps retain abundant spatial detail information and texture characteristics. During these early stages, the network primarily focuses on fundamental visual elements such as edges, corners, and local texture patterns, with high activation regions typically corresponding to object boundaries and surface texture variations. Particularly in underwater environments, these shallow features effectively capture boundary information between debris targets and complex seabed backgrounds, establishing a solid foundation for subsequent target segmentation and localization. Notably, compared to traditional network architectures, SPyramidLightNet demonstrates superior feature fidelity in shallow stages, primarily attributed to the dynamic kernel selection mechanism in the SMAB module, enabling the network to adaptively adjust feature extraction strategies based on local input characteristics.

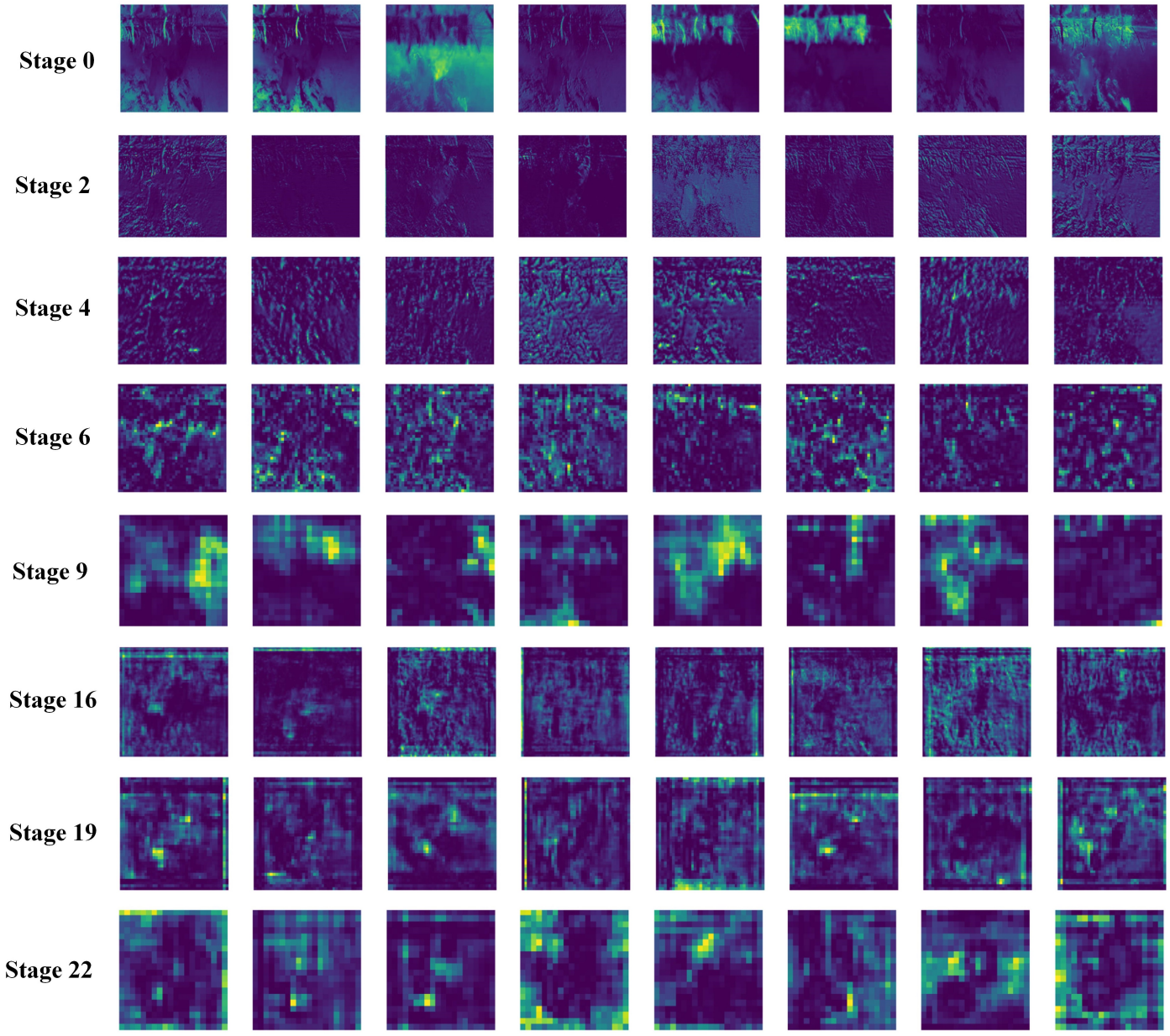



**Figure 6.** Feature Map Visualization at Different Stages.

The evolutionary process of middle-layer features (Stage 4 through Stage 9) showcases SPyramidLightNet's distinctive multi-scale feature fusion capabilities. During this phase, feature maps begin exhibiting more abstract representation patterns, with spatial resolution gradually decreasing while semantic information continuously enriches. Stage 6 feature maps display evident target focusing effects, with high activation regions concentrating toward potential target locations, indicating the network's preliminary capability for candidate region identification. By Stage 9, feature maps present more pronounced block-wise activation patterns, reflecting the effectiveness of the SPyramidConv module in multi-scale information aggregation. Through parameter-shared multi-dilation-rate convolution architecture, the network maintains feature extraction consistency across different receptive field scales, enabling middle-layer features to achieve stronger semantic expression while preserving spatial position information.

The deep feature visualization results (Stage 16 through Stage 22) thoroughly validate SPyramidLightNet's superior performance in high-level semantic understanding. In these deep stages, feature maps exhibit highly abstract activation patterns where spatial detail information is somewhat reduced but semantic representation capability significantly enhanced. Feature maps from Stage 16 and Stage 19 display relatively stable activation distributions, indicating the network has formed stable internal representations for different debris categories. This stability holds critical importance for improving detection accuracy and reducing false positives, particularly in complex underwater environments where stable deep feature representations help distinguish genuine targets from environmental noise. Stage 22, as the network's deepest features, presents highly compressed yet information-rich representation patterns that directly feed into the C3GNHead detection head for final target classification and localization predictions.

Through comparative analysis of activation patterns across different layer feature maps, the advantages of SPyramidLightNet's hierarchical feature learning become evident. High fidelity in shallow features ensures effective preservation of detail information, multi-scale fusion capability in middle layers enhances network adaptability to targets of varying sizes, and semantic abstraction ability in deep features provides robust support for accurate target recognition. This feature learning paradigm progressing from local to global and concrete to abstract highly aligns with the intrinsic requirements of underwater debris detection tasks, effectively addressing challenges such as diverse target morphologies, dramatic scale variations, and complex backgrounds in underwater environments.

#### 4.2.2. Heatmap Visualization Analysis

To thoroughly investigate SPyramidLightNet's feature extraction and target localization mechanisms in complex underwater environments, we employ Gradient-weighted Class Activation Mapping (Grad-CAM) technology for visualizing the network's attention distribution. By generating high-resolution heatmaps, we can intuitively demonstrate the network's feature attention patterns when processing various types of underwater debris, validating the effectiveness of the proposed module designs. Figure 7 presents SPyramidLightNet's heatmap results in typical underwater debris detection scenarios.

Analysis of the heatmaps in Figure 7 reveals SPyramidLightNet's exceptional target localization and feature extraction capabilities in complex underwater environments. First, regarding target boundary localization, the network-generated heatmaps exhibit highly precise and focused activation patterns. For regularly shaped debris targets (such as plastic bottles and metal cans), heatmaps accurately cover the complete object contours, with activation regions closely matching actual target boundaries. Notably, for irregularly shaped debris (including plastic bags, fishing nets, and gloves), the network equally identifies complete target shapes, with activation areas tightly conforming to actual boundaries,

demonstrating the effectiveness of multi-scale dynamic convolutions in the SMAB module when handling complex-shaped targets.

Second, SPyramidLightNet demonstrates robust discriminative capability in background interference suppression. Common underwater elements such as seabed rocks, sand, and aquatic vegetation produce minimal activation responses in the heatmaps, with the network accurately distinguishing targets from background environments. This characteristic proves crucial for reducing false detections, particularly in seabed environments where debris often shares color and texture similarities with surroundings. The multi-dilation-rate convolution design in the SPyramidConv module plays a pivotal role here, enabling the network to better capture subtle differences between targets and backgrounds through feature fusion across different receptive fields.

Regarding multi-target scene processing capabilities, heatmap results demonstrate the network's excellent multi-scale detection performance. When images contain multiple debris targets of varying scales simultaneously, the network generates relatively balanced activation intensities for each target, avoiding the phenomenon of large targets suppressing smaller ones. This balanced activation distribution benefits from C3GNHead's shared convolution mechanism and scale-adaptive adjustment strategy, ensuring adequate network attention for targets across all scales.

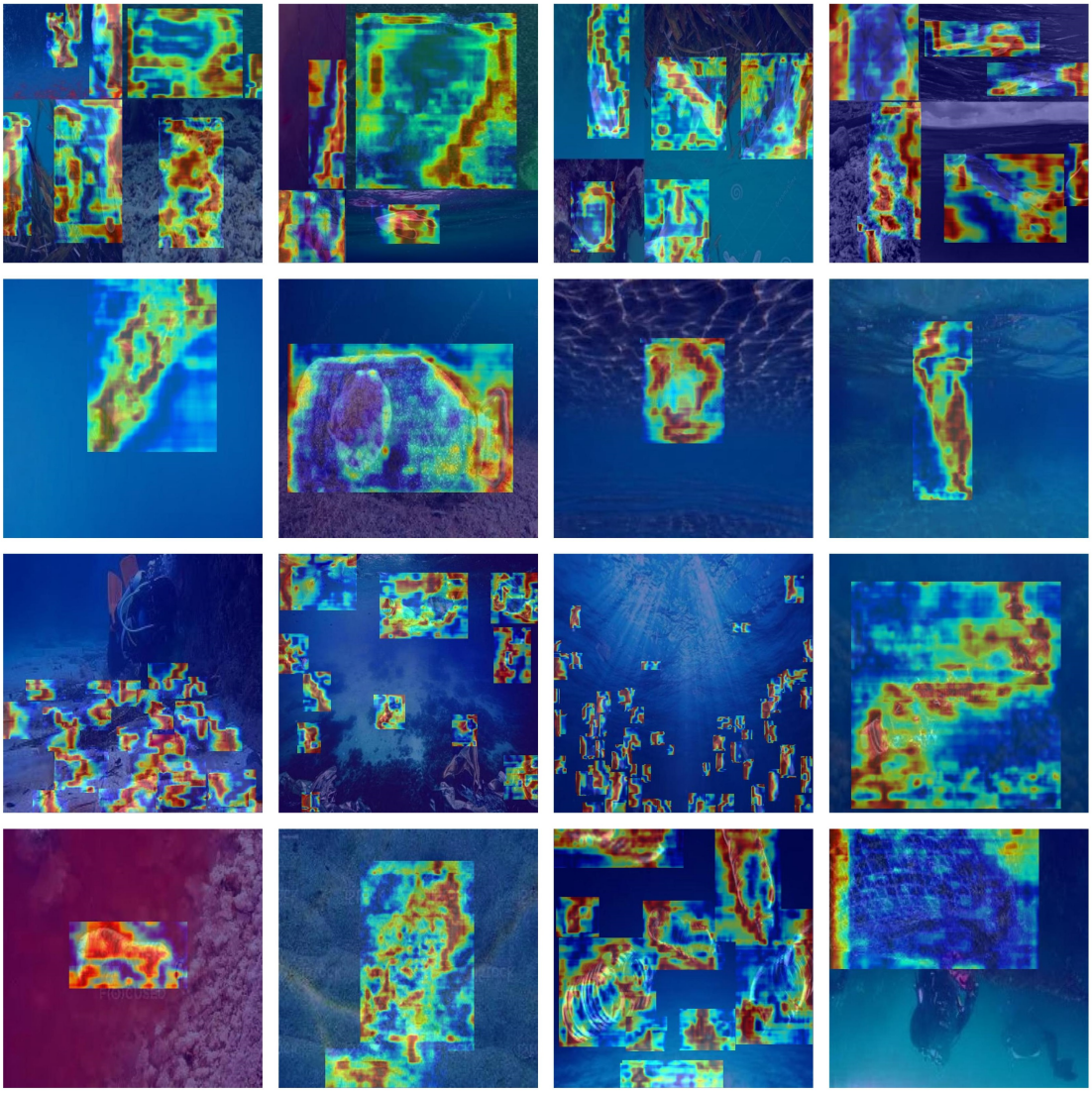


**Figure 7.** Heatmaps.

The heatmap visualization analysis comprehensively validates the design rationality and synergistic effects of SPyramidLightNet's core modules. The SMAB module enhances target boundary localization precision through multi-scale dynamic convolutions; the SPyramidConv module strengthens background interference suppression via multi-dilation-rate feature fusion; and C3GNHead ensures balanced multi-scale detection through its sharing mechanism. This intuitive visualization verification not only enhances algorithm interpretability but also provides confidence assurance for practical applications of underwater debris detection algorithms.

### *4.3. Comparative Experiments*

To comprehensively validate SPyramidLightNet's effectiveness, we conducted extensive comparative experiments against various mainstream object detection algorithms on the underwater debris dataset. The experiments selected representative algorithms covering different design paradigms, including the two-stage detector Faster-RCNN, single-stage detector SSD, Transformer-based RT-DETR, and multiple YOLO series versions (YOLOv3 through YOLOv11). All algorithms were trained and tested under identical experimental conditions, employing unified data preprocessing strategies and evaluation criteria to ensure fairness and credibility of comparative results. Table 1 presents quantitative comparison results across algorithms on the underwater debris dataset.

From a detection accuracy perspective, SPyramidLightNet demonstrates outstanding performance on key evaluation metrics. For the more stringent mAP@0.5:0.95 metric, SPyramidLightNet achieves 0.416, significantly surpassing all comparison algorithms and improving by 1.1 percentage points over the closest performer, YOLOv6. This result thoroughly validates the proposed algorithm's detection accuracy advantages across different IoU thresholds, particularly exhibiting clear superiority in high-precision application scenarios. For mAP@0.5, SPyramidLightNet reaches 0.659, leading among YOLO-series algorithms and confirming its algorithmic effectiveness under a standard detection accuracy evaluation. Although SPyramidLightNet's Recall metric of 0.616 falls slightly below RT-DETR and SSD, this performance remains aligned with practical application requirements considering its significant lightweight advantages.

Examining computational efficiency dimensions, SPyramidLightNet exhibits exceptional lightweight characteristics. Regarding computational complexity, SPyramidLightNet requires only 5.2 GFLOPs, the lowest among all comparison algorithms, representing a 17.5% reduction from baseline YOLOv11 and a remarkable 98.6% decrease compared to traditional Faster-RCNN, fully demonstrating the effectiveness of the proposed lightweight design. For inference speed, SPyramidLightNet achieves 384 FPS, second only to YOLOv11's 357 FPS and far exceeding other comparison algorithms, satisfying stringent real-time detection requirements. This superior inference efficiency primarily stems from SMAB module's parameter optimization design, C3GNHead's shared convolution mechanism, and SPyramidConv's efficient feature aggregation strategy.

Comprehensive analysis of the accuracy–efficiency balance indicates that SPyramidLightNet achieves optimal trade-offs between detection performance and computational overhead. Compared to traditional high-accuracy algorithms like Faster-RCNN, SPyramidLightNet maintains comparable detection accuracy while dramatically reducing computational complexity. Against similar lightweight algorithms in the YOLO series, SPyramidLightNet achieves performance improvements on critical accuracy metrics while significantly reducing parameters and computations. This optimized accuracy–efficiency balance provides crucial technical support for practical deployment, particularly in resource-constrained underwater application environments.

Furthermore, analyzing algorithmic generalization capability, SPyramidLightNet's stable performance in complex underwater environments validates its design robustness. Unlike other algorithms that may experience performance fluctuations when facing underwater-specific challenges (such as light attenuation, color distortion, and complex backgrounds), SPyramidLightNet maintains consistent detection performance across various complex scenarios through multi-scale feature fusion and dynamic attention mechanisms, providing a reliable solution for underwater debris detection tasks.

**Table 1.** Comparative Experiments of Different Models on the Underwater Debris Dataset.

| Algorithm | Recall | mAP@0.5 | mAP@0.5:0.95 | FPS | GFLOPs |
|---|---|---|---|---|---|
| Faster-RCNN | 0.416 | 0.485 | 0.299 | 2.1 | 370.2 |
| RT-DETR | 0.676 | 0.643 | 0.407 | 223 | 57 |
| SSD | 0.675 | 0.683 | 0.331 | 51.1 | 275.4 |
| YOLOv3 | 0.647 | 0.676 | 0.352 | 356.2 | 19 |
| YOLOv5 | 0.651 | 0.679 | 0.395 | 250.4 | 24.1 |
| YOLOv6 | 0.672 | 0.682 | 0.423 | 284.6 | 11.9 |
| YOLOv8 | 0.612 | 0.658 | 0.403 | 293.8 | 8.2 |
| YOLOv10 | 0.599 | 0.636 | 0.382 | 244.8 | 8.4 |
| YOLOv11 | 0.608 | 0.646 | 0.405 | 357 | 6.3 |
| SPyramidLightNet | 0.616 | 0.659 | 0.416 | 384 | 5.2 |

### *4.4. Ablation Study*

To thoroughly verify the effectiveness of SPyramidLightNet's core modules and quantify their specific contributions to lightweight design, we conducted systematic ablation experiments. Using YOLOv11 as the baseline model, we progressively introduced the three core modules—SMAB, C3GNHead, and SPyramidConv—comprehensively evaluating each component's independent contribution and synergistic effects on detection performance and model lightweighting. Tables 2 and 3 present quantitative comparison results of different module combinations in terms of detection performance and computational efficiency, respectively.

From a lightweight effectiveness perspective, each module demonstrates significant parameter optimization capability. The introduction of the SMAB module achieves the most prominent parameter compression, reducing model parameters from the baseline's 2,624,080 to 2,362,072—a 10.0% reduction—while decreasing computational complexity from 6.3 GFLOPs to 5.9 GFLOPs, a 6.3% reduction. This remarkable lightweight effect primarily stems from the effective parameter-sharing design of SMAB's dynamic kernel selection mechanism and split–merge strategy. The C3GNHead module similarly achieves substantial model compression, reducing parameters to 2,431,907 (7.3% reduction) and computational complexity to 5.7 GFLOPs (9.5% reduction). Through shared convolution mechanisms and GroupNorm optimization strategies, this module significantly reduces computational overhead while maintaining detection head functional integrity.

Although the SPyramidConv module slightly increases parameters to 2,740,205 when used independently, its parameter-shared multi-dilation-rate convolution design achieves more efficient feature extraction, establishing a solid foundation for overall lightweighting. More importantly, when the three modules work synergistically, they exhibit exceptional lightweight synergy effects, with final model parameters reduced to 2,313,114—an 11.8% reduction from baseline—and computational complexity decreased to 5.2 GFLOPs, a 17.5% reduction. This super-linear lightweight effect validates the high consistency of design philosophy across modules and effective fusion of optimization strategies.

Examining the balance between detection performance and lightweighting, ablation results thoroughly demonstrate the effectiveness of the proposed lightweight strategies.

The SPyramidConv module produces the most significant positive impact on detection accuracy while achieving lightweighting, improving mAP@0.5:0.95 from baseline's 0.405 to 0.419—a 3.5% increase. This result indicates that carefully designed parameter-sharing mechanisms not only reduce model complexity but also enhance feature representation capability, achieving dual objectives of lightweighting and performance improvement. The SMAB module maintains mAP@0.5:0.95 at 0.407 while substantially reducing parameters, demonstrating the robustness of dynamic convolution design in lightweight scenarios.

The complete model's performance further validates the success of lightweight design, while achieving significant parameter compression and computational optimization, SPyramidLightNet achieves an mAP@0.5:0.95 of 0.416, not only surpassing the baseline model but also maintaining excellent detection accuracy while dramatically reducing computational costs. This optimized performance-efficiency balance fully reflects the scientific validity and practicality of each module's lightweight design, providing an ideal solution for underwater object detection in resource-constrained environments.

**Table 2.** Detection Performance Comparison of Optimized Module Ablation Studies on the Underwater Debris Dataset.

| Algorithm | SMAB | C3GN | SPyramidConv | Precision/% | Recall/% | mAP@0.5/% | mAP@0.5:0.95/% |
|---|---|---|---|---|---|---|---|
| 1 | | | | 0.727 | 0.608 | 0.646 | 0.405 |
| 2 | ✓ | | | 0.741 | 0.61 | 0.651 | 0.407 |
| 3 | | ✓ | | 0.739 | 0.603 | 0.648 | 0.403 |
| 4 | | | ✓ | 0.753 | 0.621 | 0.668 | 0.419 |
| 5 | ✓ | ✓ | ✓ | 0.732 | 0.616 | 0.659 | 0.416 |

**Table 3.** Comparison of Parameter Size and Computational Complexity in Ablation Studies of Optimized Modules on the Underwater Debris Dataset.

| Algorithm | SMAB | C3GN | SPyramidConv | Parameters | GLOPS |
|---|---|---|---|---|---|
| 1 | | | | 2,624,080 | 6.3 |
| 2 | ✓ | | | 2,362,072 | 5.9 |
| 3 | | ✓ | | 2,431,907 | 5.7 |
| 4 | | | ✓ | 2,740,205 | 6.5 |
| 5 | ✓ | ✓ | ✓ | 2,313,114 | 5.2 |

### *4.5. Detection Result Visualization*

To intuitively demonstrate SPyramidLightNet's detection performance in complex underwater environments, Figure 8 presents typical detection results across different scenarios. These results encompass the main debris categories in the dataset, thoroughly validating the proposed algorithm's detection capability and robustness in diverse underwater environments. The visualization results clearly show that SPyramidLightNet can accurately identify and localize underwater debris targets under various challenging conditions.

From a detection accuracy perspective, SPyramidLightNet exhibits excellent target recognition capability. Most detection results achieve high confidence scores above 0.8, with common debris categories such as plastic bags (pbag), masks, and plastic bottles (pbottle) generally exceeding 0.85 confidence, fully demonstrating the network's high-confidence recognition capability for these targets. Particularly when processing geometrically regular targets like plastic bottles, the algorithm achieves high-confidence detection of 0.89, proving the effectiveness of the SMAB module in handling targets with distinct geometric features.

Examining target localization precision, the high alignment between detection boxes and actual target boundaries validates the algorithm's excellent spatial localization capability. Whether for regular-shaped tires and plastic bottles or irregular-shaped plastic bags

and fishing nets, detection boxes closely conform to actual target contours with extremely high boundary localization accuracy. This outstanding performance primarily results from the SPyramidConv module's multi-scale feature fusion capability, enabling the network to accurately capture boundary information of differently shaped targets.

Regarding complex environment adaptability, SPyramidLightNet demonstrates robust performance. The visualization results show that the algorithm maintains stable detection performance under various underwater environmental conditions, including dimly lit deep-water environments, turbid nearshore areas, and complex seabed terrains. Particularly when processing targets highly integrated with seabed environments, such as tires partially buried in sand or miscellaneous debris similar to rocky backgrounds, the network still achieves accurate recognition and precise localization, reflecting C3GNHead detection head's excellent discriminative capability under complex background interference.

Multi-target detection capability is equally well-validated. In complex scenes containing multiple debris targets, such as images with both fishing nets and other debris, SPyramidLightNet can simultaneously detect all targets accurately, avoiding mutual interference or omissions between targets. This parallel multi-target detection capability holds significant importance for practical marine cleanup applications, substantially improving debris identification efficiency and coverage.

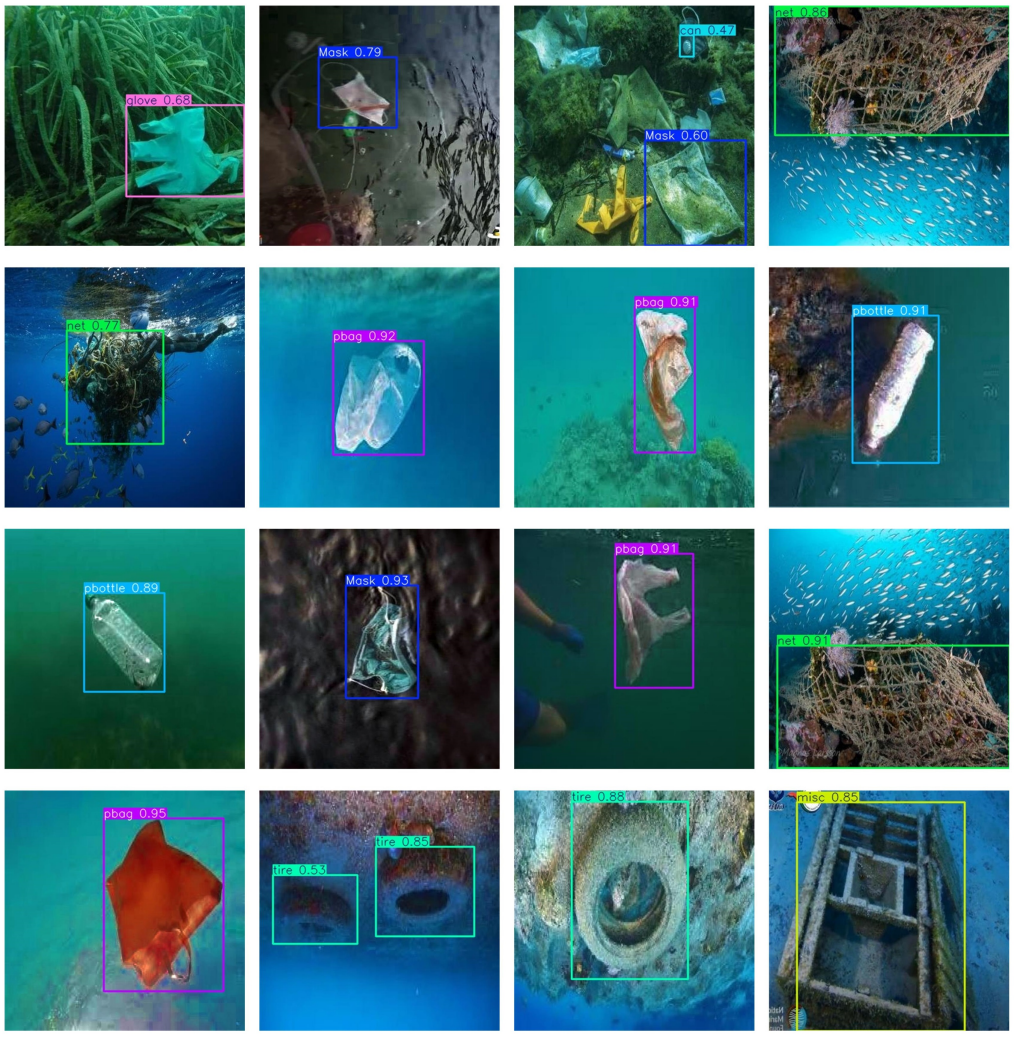



**Figure 8.** Detection Results.

Overall, the detection result visualization in Figure 8 thoroughly validates SPyramid-LightNet's excellent performance in underwater debris detection tasks. High-confidence

detection results, precise target localization, robust environmental adaptability, and stable multi-target detection performance collectively demonstrate the proposed algorithm's reliability and practical value in real marine environment applications, providing strong technical support for automated underwater debris identification and cleanup.

### 4.6. Robustness Evaluation Under Degraded Underwater Imaging

One of the most critical challenges in underwater computer vision lies in the severe degradation of visual data caused by environmental and sensor-induced artifacts. Typical distortions such as random noise, motion blur, and spectral color shifts significantly reduce image quality and can drastically impair detection performance if not adequately addressed. A second limitation frequently encountered in underwater research is the reliance on a single benchmark dataset, which, while useful for controlled evaluation, does not fully capture the variability of real-world conditions and thus limits conclusions regarding model universality.

To address these two concerns simultaneously, we extended our evaluation of SPyramidLightNet to include a diverse set of degraded underwater images that deviate substantially from the distribution of the primary dataset. The degraded images were sourced from a different dataset to ensure an independent evaluation setting. These additional samples encompass noisy sonar imagery, blurred optical captures, and strongly color-biased frames, thereby approximating the adverse conditions encountered in practical deployments. As depicted in Figure 9, SPyramidLightNet exhibits remarkable resilience: even under extreme degradations, the model consistently localizes underwater debris with well-aligned bounding boxes and high confidence scores. This qualitative robustness indicates that the network architecture shows promising generalization potential beyond the primary training dataset, demonstrating resilience to certain types of image degradation commonly encountered in underwater environments.

Beyond qualitative evidence, we provide quantitative ablation results in Table 4. The analysis demonstrates that each proposed component—SMAB, C3GNHead, and SPyramidConv—contributes measurable improvements under degraded conditions, culminating in the full SPyramidLightNet achieving a precision of 0.918, recall of 0.916, and mAP@0.5 of 0.919. These metrics surpass the YOLOv11 baseline across all evaluation dimensions, even when the input images diverge significantly from the training distribution. The consistency of these gains underscores both the robustness of the network to visual artifacts and its universality beyond a single curated dataset.

Taken together, these experiments provide compelling evidence that SPyramidLightNet is not only lightweight and accurate but also inherently resilient to noise, blur, and color distortion. More importantly, by demonstrating strong performance on images outside the training distribution, the model exhibits tangible generalization potential, thereby addressing the limitation of single-dataset evaluation. As future work, we aim to expand this validation to include multiple publicly available underwater datasets, further consolidating the universality of the proposed approach.

**Table 4.** Ablation results under degraded conditions.

| Model Configuration | Precision (P) | Recall (R) | mAP@0.5 | mAP@0.5:0.95 |
|---|---|---|---|---|
| Baseline (YOLOv11) | 0.887 | 0.846 | 0.892 | 0.631 |
| SMAB | 0.891 | 0.861 | 0.901 | 0.637 |
| C3GNHead | 0.899 | 0.873 | 0.908 | 0.639 |
| SPyramidConv | 0.903 | 0.881 | 0.912 | 0.636 |
| SPyramidLightNet (All Modules) | 0.918 | 0.916 | 0.919 | 0.642 |

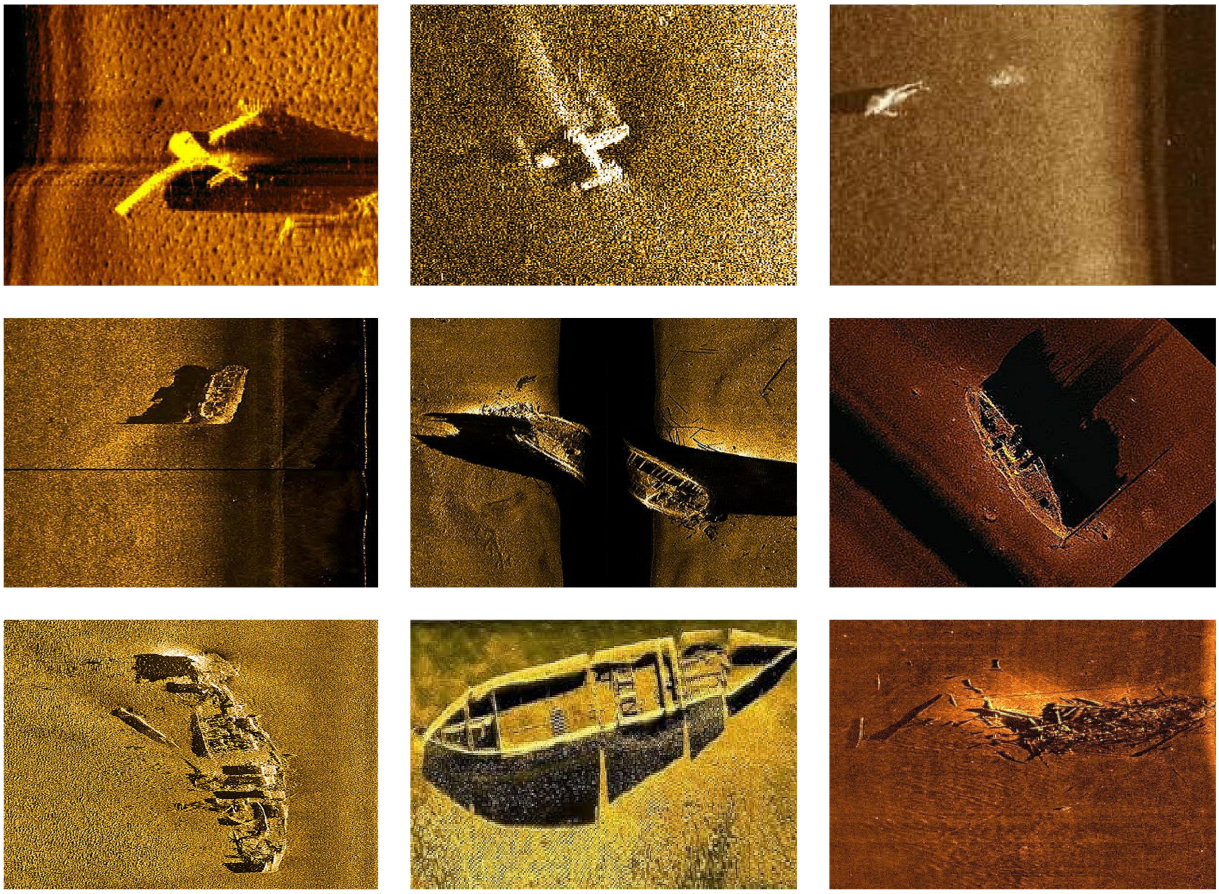


**Figure 9.** Detection results on degraded underwater images.

## 5. Conclusions

This paper addresses the dual challenges of excessive model complexity and insufficient detection accuracy in underwater debris detection by proposing a novel lightweight object detection network, SPyramidLightNet. By integrating three core innovative modules, this network successfully achieves optimal balance between detection performance and computational efficiency, providing an excellent solution for real-time underwater object detection in resource-constrained environments. The main technical contributions of SPyramidLightNet are manifested in three aspects. First, the Split–Merge Attention Block (SMAB) enhances multi-scale feature extraction capability while significantly reducing parameters by 10.0% through dynamic kernel selection mechanisms and split–merge strategies. Second, the C3 GroupNorm Detection Head (C3GNHead) achieves lightweight improvement of the detection head with 7.3% parameter compression by employing shared convolution mechanisms and GroupNorm optimization strategies. Finally, the Shared Pyramid Convolution (SPyramidConv) overcomes information loss issues of traditional pooling operations through parameter-shared multi-dilation-rate convolution design, establishing a foundation for refined feature extraction. Experimental results thoroughly validate the effectiveness and advancement of the proposed algorithm. On the underwater debris dataset, SPyramidLightNet achieves 0.416 mAP@0.5:0.95, significantly surpassing all comparison algorithms including the YOLO series. In terms of lightweighting, it reduces parameters by 11.8% and computational complexity by 17.5% compared to the baseline YOLOv11, while maintaining a high inference speed of 384 FPS, satisfying stringent real-time detection requirements. Ablation experiments further confirm the independent contributions and synergistic effects of each core module, while heatmap visualization analysis and detection results demonstrate the algorithm's excellent performance in complex underwater environ-

ments. This research provides important theoretical guidance and practical reference for lightweight object detection algorithm design, with the proposed parameter sharing, dynamic weight allocation, and multi-scale feature fusion strategies offering valuable insights for related fields. Meanwhile, SPyramidLightNet's successful application in underwater debris detection tasks provides strong technical support for marine environmental protection and automated cleanup technology development, possessing significant social value and application prospects.

Despite the promising results, several aspects warrant further investigation. Future work will include quantitative analyses of environmental factors (e.g., depth, turbidity, illumination) to enhance robustness assessment, as well as real-time deployment and validation on AUVs and underwater platforms. Power consumption, a key factor in marine applications, will be systematically evaluated. To strengthen generalization, we will extend testing to multiple public datasets. Moreover, we plan to incorporate composite metrics that jointly reflect detection accuracy and computational efficiency for a more holistic evaluation.

**Author Contributions:** Conceptualization, Y.L. and O.E.; Methodology, Y.L.; Software, Y.L.; Validation, Y.L.; Formal analysis, Y.L.; Investigation, Y.L.; Resources, O.E.; Data curation, Y.L.; Writing—original draft, Y.L.; Writing—review & editing, O.E.; Visualization, Y.L.; Supervision, O.E.; Project administration, O.E.; Funding acquisition, O.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

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

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

## References

1. Jambeck, J.R.; Geyer, R.; Wilcox, C.; Siegler, T.R.; Perryman, M.; Andrady, A.; Law, K.L. Plastic waste inputs from land into the ocean. *Science* **2015**, *347*, 768–771. [CrossRef]
2. Woodall, L.C.; Sanchez-Vidal, A.; Canals, M.; Paterson, G.L.; Coppock, R.; Sleight, V.; Thompson, R.C. The deep sea is a major sink for microplastic debris. *R. Soc. Open Sci.* **2014**, *1*, 140317. [CrossRef] [PubMed]
3. Wright, S.L.; Thompson, R.C.; Galloway, T.S. The physical impacts of microplastics on marine organisms: A review. *Environ. Pollut.* **2013**, *178*, 483–492. [CrossRef] [PubMed]
4. Fulton, M.; Hong, J.; Islam, M.J.; Sattar, J. Robotic detection of marine litter using deep visual detection models. In Proceedings of the IEEE International Conference on Robotics and Automation 2019, Montreal, QC, Canada, 20–24 May 2019; pp. 5752–5758.
5. Li, X.; Zhao, Y.; Wang, J.; Chen, L. Real-time underwater object detection technology for complex underwater environments based on deep learning. *Neurocomputing* **2024**, *578*, 127394.
6. Er, M.J.; Chen, J.; Zhang, Y.; Gao, W. Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. *Sensors* **2023**, *23*, 1990. [CrossRef]
7. Xu, S.; Zhang, M. A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing* **2023**, *527*, 204–232. [CrossRef]
8. Akkaynak, D.; Treibitz, T. Sea-thru: A method for removing water from underwater images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1682–1691.
9. Watson, J.; Zielinski, A. Underwater image enhancement using polarimetric filtering. *IEEE Trans. Image Process.* **2013**, *22*, 1934–1946.
10. Islam, M.J.; Edge, C.; Xiao, Y.; Luo, P.; Ahmed, M.; Morse, C.; Sattar, J. Semantic segmentation of underwater imagery: Dataset and benchmark. In Proceedings of the 2020 IEEE/CVF International Conference on Computer Vision, Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 1769–1778.

11. Fan, Y.; Liu, J.; Zhang, M.; Wang, L. A small underwater object detection model with enhanced feature extraction and fusion. *Sci. Rep.* **2024**, *15*, 1456.
12. Palomeras, N.; Ridao, P.; Ribas, D.; Vallicrosa, G.; Mallios, A.; Camilli, R.; Gracias, N. Autonomous underwater vehicle trajectory planning for real-time photogrammetric survey. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3132–3139.
13. Moniruzzaman, M.; Islam, S.M.S.; Bennamoun, M.; Lavery, P. Deep learning on underwater marine object detection: A survey. In *Advanced Concepts for Intelligent Vision Systems*; Springer: Cham, Swizterland, 2017; pp. 150–160.
14. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
15. Zhuang, Y.; Li, J.; Wang, H.; Chen, S.; Liu, M.; Tang, S.; Liu, C. A deep learning framework based on structured space model for detecting small objects in complex underwater environments. *Commun. Eng.* **2025**, *6*, 47. [CrossRef]
16. Chen, L.; Liu, Z.; Jiang, Z.; Wang, S.; Dong, J. YOLO-UOD: An underwater small object detector via improved efficient layer aggregation network. *IET Image Process.* **2024**, *18*, 1564–1576. [CrossRef]
17. Feng, J.; Jin, T. CEH-YOLO: A composite enhanced YOLO-based model for underwater object detection. *Ecol. Inform.* **2024**, *82*, 102758. [CrossRef]
18. Yeh, C.H.; Lin, C.H.; Kang, L.W.; Huang, C.H.; Lin, M.H.; Chang, C.Y.; Wang, C.C. Lightweight deep neural network for joint learning of underwater object detection and color conversion. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6129–6143. [CrossRef]
19. Chen, Y.; Wang, H.; Li, S.; Zhang, M.; Liu, K.; Zhou, L. A comprehensive review of model compression techniques in machine learning. *Appl. Intell.* **2024**, *55*, 1247–1285.
20. Han, S.; Pool, J.; Tran, J.; Dally, W.J. Learning both weights and connections for efficient neural networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1135–1143.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
22. Zhou, Y.; Chen, L.; Wang, H.; Zhang, M.; Liu, S. Underwater small object detection via automatic encoding pyramid with neighborhood information. *J. Ocean Eng. Mar. Energy* **2025**, *11*, 405–420.
23. Li, J.; Wang, S.; Chen, H.; Zhang, L.; Liu, M. A survey of model compression strategies for object detection. *Multimed. Tools Appl.* **2023**, *82*, 45123–45158.
24. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2019**, *128*, 261–318. [CrossRef]
25. Chen, Y.; Liu, S.; Wang, H.; Zhang, M.; Li, J. Optimizing Edge AI: A comprehensive survey on data, model, and system strategies. *arXiv* **2024**, arXiv:2501.03265.
26. Yi, Y.; Zhang, L.; Wang, S.; Liu, H. Efficient neural networks for edge devices. *Comput. Netw.* **2021**, *195*, 108157.
27. Kamath, V.; Renuka, A. A comprehensive survey of deep learning-based lightweight object detection models for edge devices. *Artif. Intell. Rev.* **2024**, *57*, 242. [CrossRef]
28. Ahmed, S.; Khan, M.; Li, H.; Wang, J.; Zhang, L. Lightweight Deep Learning for Resource-Constrained Environments: A Survey. *ACM Comput. Surv.* **2024**, *56*, 1–42. [CrossRef]
29. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
30. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
31. Tan, M.; Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 2019 International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
32. Koonce, B. MobileNetV3. In *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: Berkeley, CA, USA, 2021; pp. 125–144.
33. Zhang, X.; Zhou, X.; Lin, M. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
34. Ma, N.; Zhang, X.; Zheng, H.T. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the 2018 Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
35. Koonce, B. EfficientNet. In *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: Berkeley, CA, USA, 2021; pp. 109–123.
36. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; PMLR, 2021; pp. 10096–10106.
37. Fan, B.; Chen, Y.; Qu, J.; Chai, Y.; Xiao, C.; Huang, P. FFBNet: Lightweight backbone for object detection based feature fusion block. In Proceedings of the IEEE International Conference on Image Processing, Taipei, Taiwan, 22–25 September 2019; pp. 3920–3924.

38. Liu, W.; Liao, S.; Ren, W.; Hu, W.; Yu, Y. High-level semantic feature detection: A new perspective for pedestrian detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5187–5196.
39. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
40. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
41. Wang, A.; Chen, H.; Liu, L. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011.
42. Liu, W.; Anguelov, D.; Erhan, D. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016, Proceedings, Part I 14; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37.
43. Xiong, Y.; Liu, H.; Gupta, S. Mobiledets: Searching for object detection architectures for mobile accelerators. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3825–3834.
44. Lin, T.Y.; Goyal, P.; Girshick, R. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
45. Lin, T.Y.; Dollár, P.; Girshick, R. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
46. Liu, S.; Qi, L.; Qin, H. Path aggregation network for instance segmentation. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
47. Chen, J.; Mai, H.S.; Luo, L. 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; pp. 699–703.
48. Zhang, T.; Jin, B.; Jia, W. An anchor-free object detector based on soften optimized bi-directional FPN. *Comput. Vis. Image Underst.* **2022**, *218*, 103410. [CrossRef]
49. Li, C.; Guo, J.; Cong, R.; Pang, Y.; Wang, B. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Trans. Image Process.* **2017**, *25*, 5664–5677. [CrossRef]
50. Drews, P.; Nascimento, E.; Moraes, F.; Botelho, S.; Campos, M. Transmission estimation in underwater single images. In Proceedings of the IEEE International Conference on Computer Vision Workshops 2013, Sydney, Australia, 2–8 December 2013; pp. 825–830.
51. Wu, J.; Liu, X.; Lu, Q. FW-GAN: Underwater image enhancement using generative adversarial network with multi-scale fusion. *Signal Process. Image Commun.* **2022**, *109*, 116855. [CrossRef]
52. Yin, X.; Chen, X.; Yang, Y. Underwater Image Enhancement Algorithm Based on UWGAN and U-Net. In *Advances in Machinery, Materials Science and Engineering Application IX*; IOS Press: Amsterdam, The Netherlands, 2023; pp. 870–878.
53. Li, C.; Yu, D.; Li, H. Underwater Image Enhancement Method Based on MuLA-GAN. In *Annual Conference of China Electrotechnical Society*; Springer Nature: Singapore, 2024; pp. 523–532.
54. Chen, Y.; Li, H.; Yuan, Q. Underwater Image Enhancement based on Improved Water-Net. In Proceedings of the 2022 IEEE International Conference on Cyborg and Bionic Systems (CBS), Wuhan, China, 24–26 March 2023; pp. 450–454.
55. Li, Z.; Zheng, B.; Chao, D. Underwater-Yolo: Underwater Object Detection Network with Dilated Deformable Convolutions and Dual-Branch Occlusion Attention Mechanism. *J. Mar. Sci. Eng.* **2024**, *12*, 2291. [CrossRef]
56. Chen, L.; Yang, Y.; Wang, Z. Underwater target detection lightweight algorithm based on multi-scale feature fusion. *J. Mar. Sci. Eng.* **2023**, *11*, 320. [CrossRef]
57. Yang, H.H.; Huang, K.C.; Chen, W.T. Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement. In In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 685–692.
58. Gong, T.; Zhang, M.; Zhou, Y. Underwater image enhancement based on color feature fusion. *Electronics* **2023**, *12*, 4999. [CrossRef]


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