

# 基于 YOLOv8 的轻量化水下目标检测算法

梁秀满, 赵佳阳, 于海峰

(华北理工大学 电气工程学院, 河北 唐山 063210)

**摘要:** 针对复杂水下环境导致水下目标检测时出现误检、漏检以及检测效率低等问题, 提出了一种改进 YOLOv8 模型的轻量化水下目标检测算法。首先, 为了改善颈部网络特征融合不足的问题, 将 YOLOv8 的颈部网络融合 (Bidirectional Feature Pyramid Network, BiFPN) 双向特征金字塔结构, 提高小目标层的检测效果; 其次, 针对网络中卷积模块参数量大和计算复杂度高的问题, 设计了一种自适应注意力下采样 (Adaptive-Attention Down-Sampling, AADS) 模块, 将主干网络中的卷积模块替换为 AADS 模块, 降低模型参数量和计算量; 最后, 引入大可分离核注意力机制 (Large Separable Kernel Attention, LSKA), 强化特征提取能力, 使模型能够更精确地关注重要信息, 提高目标检测精度。将改进的网络在水下目标检测数据集上进行实验, 改进后的算法与 YOLOv8 相比, 平均检测精度提升了 1.4%, 模型计算复杂度降低了 15.9%, 模型参数量减少了 43.3%, 使检测精度和检测速度之间达到了很好的平衡。

**关键词:** YOLOv8; 水下目标检测; 大可分离核注意力机制; 轻量化; 多尺度特征融合

中图分类号: TP391.4

文献标识码: A

文章编号: 1001-8891(2024)09-1015-10

## Lightweight Underwater Target Detection Algorithm Based on YOLOv8

LIANG Xiuman, ZHAO Jiayang, YU Haifeng

(College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China)

**Abstract:** To address the problems of misdetection, omission detection, and low detection efficiency when detecting underwater targets due to the complex underwater environment, a lightweight underwater target detection algorithm with an improved YOLOv8 model is proposed. First, to ameliorate the problem of insufficient feature fusion in the neck network, the neck network of YOLOv8 is fused with a BiFPN bidirectional feature pyramid structure to improve the detection of the small target layer. Second, to address the problem of the large number of parameters of the convolution module in the network and high computational complexity, an Adaptive-Attention Down-Sampling(AADS) module is designed to replace the convolution module in the backbone network to reduce the number of model parameters and amount of computation. Finally, Large Separable Kernel Attention (LSKA) is introduced to strengthen the feature extraction capability such that the model can focus on important information more accurately and improve target detection accuracy. The experimental results show that in the underwater target detection dataset, the improved algorithm improves the average detection accuracy by 1.4%, reduces the number of model parameters by 43.3%, and reduces the computational complexity of the model by 15.9% when compared with YOLOv8. This realizes a good balance between detection accuracy and detection speed.

**Key words:** YOLOv8, underwater target detection, large separable kernel attention, lightweight, multi-scale feature fusion

## 0 引言

近年来由于海洋经济和水产养殖的不断发展, 水下目标检测方法在海洋领域得到广泛应用。水下图像

收稿日期: 2024-06-25; 修订日期: 2024-07-22.

作者简介: 梁秀满 (1973-), 女, 副教授, 主要从事图像处理与检测技术。E-mail: 13230857556@126.com.

通信作者: 于海峰 (1990-), 男, 讲师, 主要从事图像处理与深度学习。E-mail: yhf5170@163.com.

基金项目: 河北省自然科学基金资助 (D2024209006); 河北省教育厅科学研究项目资助 (QN2024147)。

目标检测对于探索和保护脆弱的生态系统至关重要。随着传感器和图像处理技术的发展,人们对于水下计算机视觉的兴趣也在逐渐激增<sup>[1]</sup>。水下图像处理在探索水下环境方面具有巨大的潜力,它已被广泛应用于各种水下场合中,例如水下机器人视觉、海洋生物监测等<sup>[2]</sup>。

水下生物环境存在光线不足、物体体积小和生物分布密集等问题<sup>[3-4]</sup>,这些问题导致水下图像质量较差,给目标检测带来了难度。水下目标检测不同于地面目标检测,首先由于水下环境比地面环境更复杂,使水下图像获取困难,水下目标数据集相对于地面目标数据集少。其次是图像的拍摄距离不同,水下拍摄距离达到百米甚至千米,相比于地面拍摄距离远近百倍甚至千倍,导致水下图像中小目标占比更多,因此对水下小目标检测的性能要求会更加严格。除了以上问题导致现有的一些目标检测算法性能较差以外,目标检测算法本身存在的检测精度低和模型结构复杂的问题使目标检测难以应用到移动设备中,这也是水下目标检测算法的难点之一。因此开发针对水下生物场景的检测算法,正逐渐成为当前研究的热门话题<sup>[5]</sup>。

随着基于深度学习的目标检测方法的发展,两阶段<sup>[6]</sup>和单阶段方法<sup>[7-8]</sup>在水下目标检测中的应用越来越多<sup>[9-11]</sup>。各种特征提取<sup>[12]</sup>、融合策略<sup>[13]</sup>和损失定义已经集成到检测模型中。

传统的两阶段方法采用区域建议模块生成候选对象,然后对候选对象进行分类和定位回归。Zeng L.<sup>[9]</sup>等将标准 Faster-RCNN<sup>[6]</sup>与对抗遮挡网络相结合,以增强对遮挡目标的鲁棒性。Song P.<sup>[10]</sup>等运用 RCNN (Region-CNN) 在低对比度和物体模糊的水下环境中解决了目标检测的不确定性建模问题。虽然基于两阶段算法的水下检测方法具有较高的精度,但采用的两阶段方法导致检测速度慢且模型尺寸较大。

与两阶段检测模型相比,单阶段检测模型将目标检测任务视为单个回归问题,以减轻计算复杂度并提高检测速度。代表算法包括 SSD (Single Shot MultiBox Detector)<sup>[7]</sup>和 YOLO (You Only Look Once) 家族<sup>[14]</sup>。YOLO 系列算法作为一种具有代表性的单阶段检测算法网络,以其精度高、检测速度快、结构灵活等优点在计算机视觉应用中得到了广泛的研究和应用。

Abdullah Al Muksit 等<sup>[15]</sup>基于 YOLOv3 提出了水下鱼类检测模型 YOLO-Fish。通过改进网络结构,解决了上采样步长问题,减少对小目标生物的误检。又通过增加空间金字塔池,增加了在动态环境中检测鱼类外观的能力。Chen L Y 等<sup>[16]</sup>对 YOLOv4 检测网络进行改进,将上采样模块替换为反卷积模块,并将深

度可分离卷积纳入网络中。不仅提升了模型的精度和检测速度,同时降低了对硬件性能的要求,提高了水下目标检测性能。Pingzhu Liu 等<sup>[17]</sup>以 YOLOv5 为基础进行网络改进,通过设计 SRC3 高效的特征提取块和 CRFPN 特征融合模块,使模型能够提取更丰富的水下目标信息,融合不同尺度的有效信息。并且引入了 EIoU 损失函数,使它们能够专注于准确探测和分类水下物体,从而解决探测过程中目标重叠和遮挡的问题。Guoyan Yu 等<sup>[18]</sup>建立了一个结合交叉转换和高效挤压激励模块的网络,提出了水下生物检测方法 U-YOLOv7,解决了水下检测中背景信息干扰多、检测速度慢的问题,实现了水下生物的自主监测和准确识别。

虽然以上工作都取得了不错的效果,但是仍然存在目标检测精度低、模型不够轻量化,应用于移动设备困难等问题,需要改进的地方仍然很多。因此本文提出了一种基于 YOLOv8 改进的轻量化水下目标检测算法,解决以上在复杂水下海洋生物检测任务中存在的诸多问题。具体改进工作有以下几点:

1) 为了实现更高水平的多尺度特征融合,加强小目标特征信息的传递,颈部网络融入加权双向特征金字塔结构 (BiFPN) 改进原来的特征融合网络。

2) 针对 YOLOv8 主干网络中的卷积模块参数量大、计算复杂度高的问题,设计了一个轻量化模块 AADS,将主干网络中的卷积模块替换为 AADS 使整体网络轻量化,并且使检测精度和速度之间达到平衡。

3) 在检测大目标的颈部引入大可分离核注意力机制 (LSKA) 提高检测精度,模型能够更精确地关注重要信息,提高模型对于复杂数据的处理能力。

## 1 YOLOv8 算法改进

### 1.1 YOLOv8 算法简介

YOLOv8 作为一种快速的单阶段目标检测方法,可以用于对象识别,图像分类和分割,目前在水下航行器探测<sup>[19]</sup>和其他应用<sup>[20]</sup>中得到了广泛的采用。YOLOv8 既能在 CPU 上运行,也能在 GPU 上运行。根据不同的使用场景将模型分为 n、s、l、m、x 共 5 种类型,其中 YOLOv8n 参数量最小且检测速度最快。

YOLOv8 由主干网络、颈部网络和头部网络组成。主干网络使用与 YOLOv5 类似的结构,保留了 CBS 卷积模块,将 YOLOv5 原始的 CSPLayer 模块修改为 C2f 模块,C2f 模块具有两个卷积的跨阶段瓶颈部分组成,这样可以更好地集成高级功能和上下文信息。主干网络的最后一层采用空间金字塔结构 SPPF

(Spatial Pyramid Pooling and Feature Fusion) 结构和多个大小池化窗口来丰富网络的特征提取。颈部网络采用 FPN (Feature Pyramid Network) + PAN (Path Aggregation Network) 实现多特征提取, 丰富对小目标的检测。其中, Upsample 为上采样操作, 用于增加特征图的分辨率。Concat 模块通过合并不同层的特征对特征图进行拼接。头部网络使用头部解耦的无锚模型来处理回归和分类任务, CIou<sup>[21]</sup>和 DFL<sup>[22]</sup>损失函数用于封闭框架预测。与基于锚点的方法相比, 无锚点方法减少了参数的数量, 对于检测形状差异较大的小目标效果更好。

**1.2 改进算法总体框架**

针对水下目标检测特征不足导致漏检和误检的问题, 采用 YOLOv8n 作为基线网络, 以实现高精度和小模型尺寸。首先在颈部融入 BiFPN 金字塔结构, 进行多尺度和跨尺度连接, 从而得到更丰富的特征融合; 其次, 在主干网络中将 CBS 卷积模块替换为 AADS 下采样模块, 减少参数数量的同时提高检测精度; 最后, 在检测大目标特征的颈部与头部网络之间引入 LSKA 注意力模块, 提高对水下关键特征信息的提取, 进一步提高检测效果, 改进的 YOLOv8 网络结构如图 1 所示。

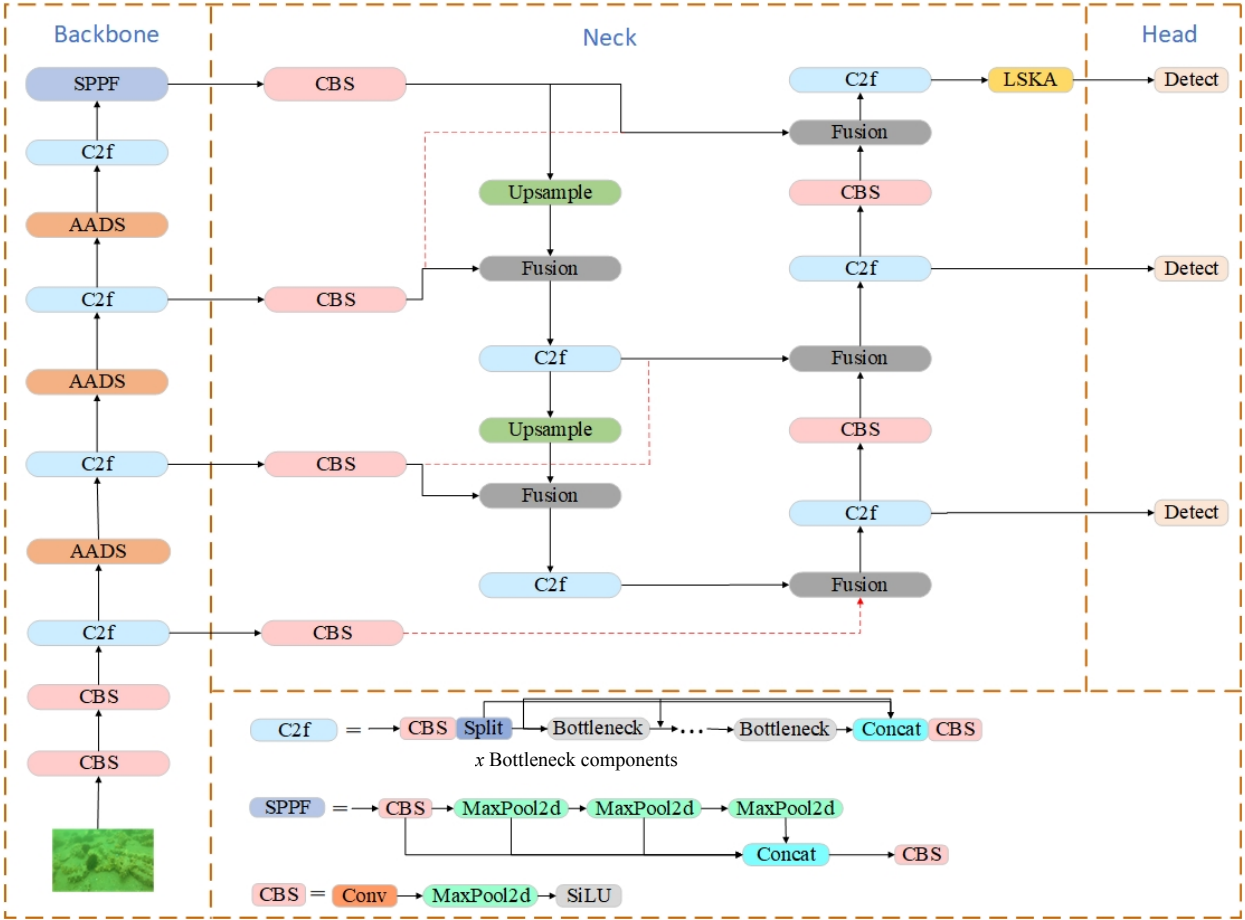


图 1 改进后的 YOLOv8 网络结构

Fig.1 Improved YOLOv8 Network Structure

1.3 双向特征金字塔

多尺度特征融合的目的是聚合不同分辨率的特征, 从而提高对不同尺度目标的检测精度。在复杂的水下场景进行生物检测, 从被遮挡的海洋生物中提取的特征信息有限, 因此会导致检测精度的降低, 为了解决这一问题, 改进的颈部网络结构在原始 YOLOv8 的 Neck 层引入双向特征金字塔网络 (BiFPN)。

传统的颈部融合结构是采用自顶向下的 FPN<sup>[23]</sup> 结构 (如图 2(a)所示), 它将自顶向下上采样纵向得到

的特征与横向连接后的特征进行多尺度融合, 把高层的特征传下来, 补充低层的语义, 进而获得高分辨率、强语义的特征。然而传统 FPN 结构在特征融合时仅考虑当前层和相邻某一层特征, 对于高层和浅层特征的融合并不充分。PANet<sup>[24]</sup> (如图 2(b)所示) 在 FPN 结构上进行改进, 在 FPN 基础上添加一个自下而上的路径聚合网络, 同时考虑了顶层的语义信息和底层的位置信息。

YOLOv8n 模型采用 FPN+PAN 的结构, 但由于

输入特性的分辨率不同, FPN+PAN 结构对融合输出特性的贡献往往是不均匀的, 不同尺度之间的特性不能得到充分利用。因此, 将颈部网络融合加权双向特征金字塔结构(图 2(c)所示)以此来加强不同尺度特征之间的融合和不同特征信息之间的传递。改进后的网络能够挖掘更深层的水下目标信息, 将深层特征信息与浅层特征信息进行信息共享, 提高模型的检测精度。

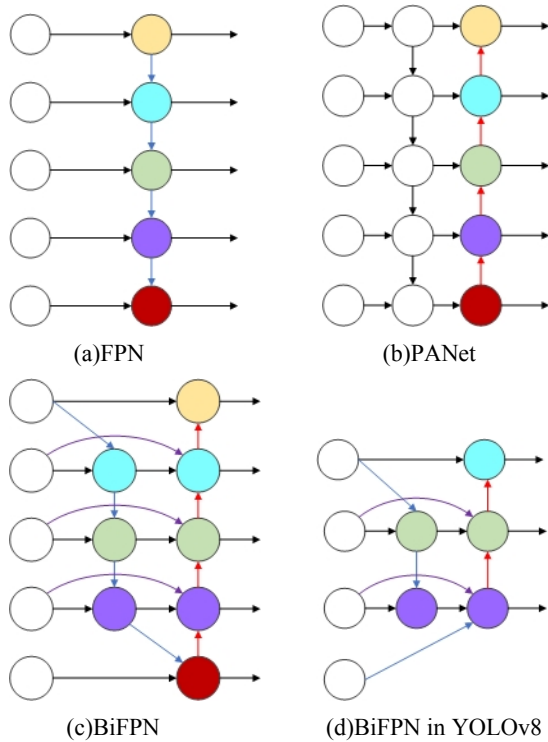


图2 特征融合网络结构

Fig.2 Structure of Feature Fusion Network

BiFPN 去除那些只有一个输入边的节点, 从而简化了双向网络结构; 为了实现低成本操作情况下融合更多尺度特征, 在原始输入节点和输出节点之间增加一条额外的边; 与 PANet 只有一条自顶向下和一条自底向上的路径不同, BiFPN 将每条双向(自顶向下和自底向上)路径视为一个特征网络层, 对同一层进行多次重复, 使不同特征融合变得更高阶。此外, 由于输入特征的分辨率不同, BiFPN 还引入了训练权重, 通过将额外的权重添加到每个输入, 并且对每个输入分配权重, 用权重来调整不同输入对输出特征的贡献程度从而对特征层进行深度融合。采用快速归一化融合来进行权重分配, 其公式如下:

$$O = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot I_i \quad (1)$$

通过在每个  $\omega_i$  使用 Relu 激活函数来保证  $\omega_i \geq 0$ , 并且为避免数值不稳定, 将  $\varepsilon$  设为极小值 0.0001, 每

个归一化权重的值都在 0~1 之间。

如图 2(d)为 BiFPN 在改进模型中的应用, 考虑到水下环境中存在小目标且难以检测的问题, 在此基础上补充低层语义信息, 即将尺寸为  $160 \times 160$  的特征图输出节点与  $80 \times 80$  的输出节点进行融合补充。在中间两层连接中各增加一个跳跃连接, 结合了本层及上下两层不同尺度的特征, 采用双向跨尺度和加权特征融合的方式实现更高层次的特征融合。有助于模型更全面地了解遮挡目标的上下文信息, 从而提高水下目标检测的准确性。

#### 1.4 自适应注意力下采样

原始 YOLOv8 采用普通卷积对主干网络卷积进行特征提取, 由于卷积模块计算参数多, 复杂的结构使得存在计算冗余, 因此设计了一种自适应注意力下采样模块。AADS 模块借鉴了 RFA 卷积<sup>[25]</sup>的思想, RFA 卷积使用动态接受场注意力和分组卷积结构, 将学习到的注意力特征图与分组卷积提取的接受场特征信息进行交互汇总, 强调了感受野内不同特征的重要性。AADS 模块将 RFA 卷积进行改进, 进一步减少卷积的参数数量和计算复杂度。

该下采样分为两个部分, 第一部分通过加权特征得到注意力特征层, 第二部分通过分组卷积对特征进行下采样, 最后将两者得到的特征在最后一维度上进行加权求和。如图 3 为 AADS 结构图。

在注意力特征层提取部分首先对输入特征图进行  $3 \times 3$  的平均池化, 池化层的作用是对特征图进行降维和局部平均, 以提取局部平均特征。然后对平均池化过的特征图进行  $1 \times 1$  的点卷积操作, 输入输出通道不变,  $1 \times 1$  点卷积层的作用是对每个像素点进行通道间的线性组合, 调整和整合特征图的信息, 重新分配特征的通道权重, 进一步提取有效特征。最后将特征图进行重新排列, 将  $2 \times 2$  的特征块重新排列在最后一个维度上, 得到 4 个权重, 再经过 Softmax 操作, 使得每个  $2 \times 2$  的特征块的权重和为 1, 从而得到注意力权重。

在下采样特征提取部分使用  $3 \times 3$ , 步幅为 2 的分组卷积, 每个组内独立进行卷积操作, 分组卷积可以减少参数数量和计算量, 同时保留卷积操作的特征提取能力。接下来对分组卷积下采样的特征重新排列卷积输出的形状, 将输出通道拆分为 4 个通道块并进行整合, 与第一部分得到的注意力权重对应。

如图 4 为普通卷积和分组卷积<sup>[26]</sup>的对比图, 此时输入特征图尺寸为  $H \times W \times C_1$ , 普通卷积核的尺寸为  $h_1 \times w_1 \times c_1$ , 共有  $c_2$  个卷积核, 此时输出特征尺寸为  $H \times W \times C_2$ 。在分组卷积运算中, 将输入特征在深度

上分成  $g$  组, 输入高度和宽度保持不变, 即将  $c_1$  个通道分成  $g$  组, 将  $c_2$  个卷积核也分为  $g$  组, 最终的输出特征尺寸和普通卷积运算输出相同, 即  $H \times W \times C_2$ 。因为将普通卷积过程拆分成了  $g$  组子运算共同运行, 参数量得到了降低。在普通卷积中, 参数量为  $h_1 \times w_1 \times c_1 \times c_2$ , 使用分组卷积后, 参数量变为:

$$h_1 \times w_1 \times \frac{c_1}{g} \times \frac{c_2}{g} \times g = h_1 \times w_1 \times c_1 \times c_2 \times \frac{1}{g} \quad (2)$$

由此可以看出, 分组卷积在分组的过程中不仅减少了网络参数量, 而且通过分组可以减少卷积的操作数量从而减少整个网络的计算复杂度。

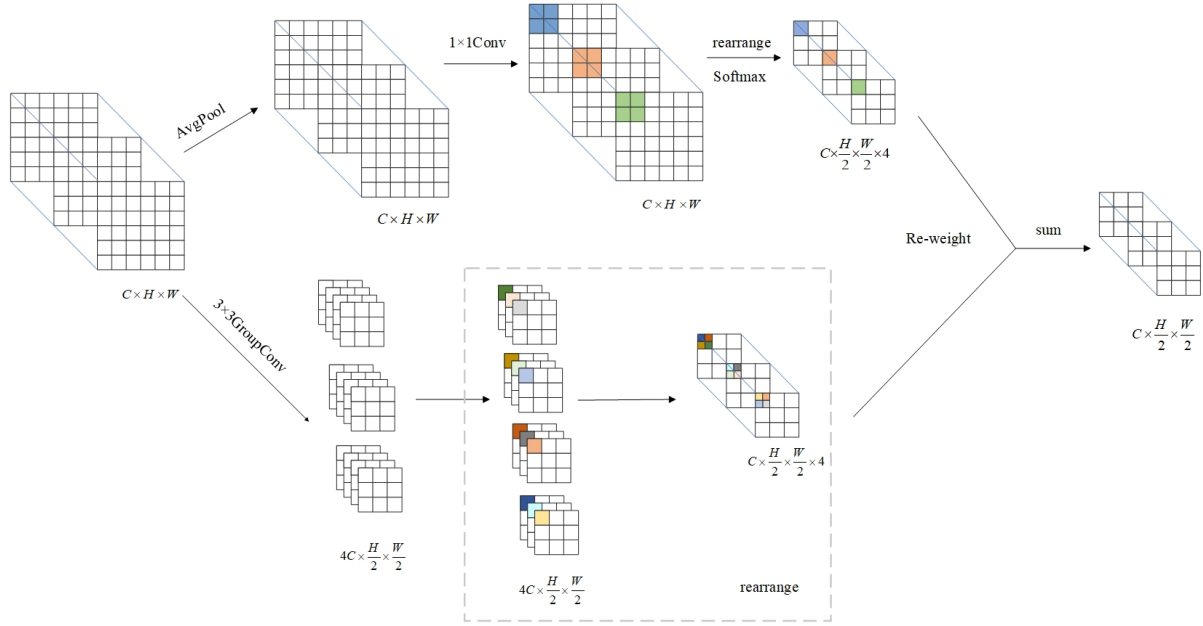


图 3 AADS 网络结构

Fig.3 AADS network structure

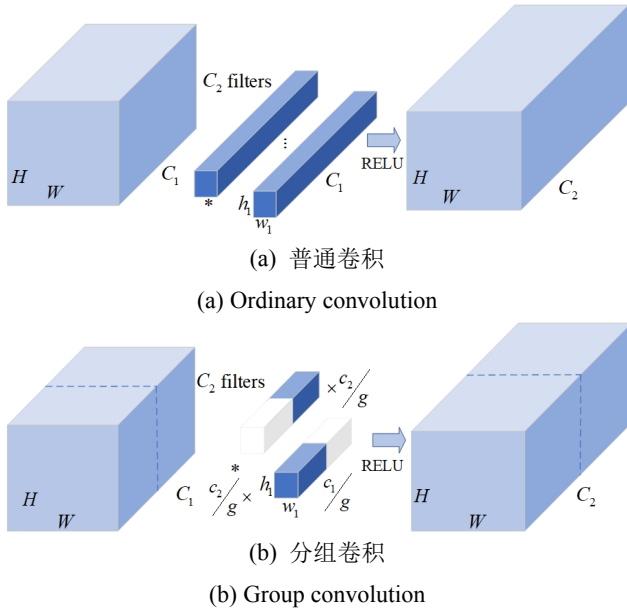


图 4 普通卷积与分组卷积对比

Fig.4 Comparison between ordinary convolution and grouped convolution

AADS 下采样模块通过在注意力特征提取阶段计算每个  $2 \times 2$  特征块的注意力权重, 自适应地关注重要特征信息, 并使用这些权重对下采样的特征图进行

加权求和, 实现了自适应权重的下采样操作。引用分组卷积作为下采样卷积使得网络计算量和参数量减少。因此将 AADS 模块替换主干网络中的卷积模块, 在减少参数量和计算复杂度的同时, 使得检测精度也得到提高。

### 1.5 大可分离核注意力机制

LSKA 注意力机制是通过 LKA<sup>[27]</sup>注意力机制改进而来, 与 LKA 不同的, LSKA 具有可以有效地捕获图像中的远程依赖关系的特性, 对于更深层网络有着更好的空间和信道适应性, 降低了计算成本。

LKA 结构图如图 5(a)所示, 它由深度卷积、深度扩展卷积和点卷积组成, LSKA 将 LKA 进行了改进, 将深度卷积和深度扩展卷积的二维核拆分为两个级联的一维可分离核, 如图 5(b)所示为 LSKA 结构图。给定一个输入特征  $F \in R^{C \times H \times W}$ , 其中  $C, H, W$  分别为通道数, 特征图高度和宽度, LSKA 的输出如下所示:

$$\bar{Z}^C = \sum_{H,W} W_{(2d-1)}^C * \left( \sum_{H,W}^C \frac{1}{l \times (2d-1)} * F^C \right) \quad (3)$$



$$Z^C = \sum_{H,W} W_{\left\lfloor \frac{k}{d} \right\rfloor \times 1}^C * \left( \sum_{H,W}^C \left\lfloor \frac{k}{d} \right\rfloor * \bar{Z}^C \right) \quad (4)$$

$$A^C = W_{1 \times 1} * Z^C \quad (5)$$

$$\bar{F}^C = A^C \otimes F^C \quad (6)$$

式中:  $d$  是膨胀率;  $*$ 和 $\otimes$ 分别表示卷积和 Hadamard 积;  $\bar{Z}^C$  表示拆分为两个级联内核大小和为 $(2d-1) \times (2d-1)$ 的深度卷积的输出, 它捕获了局部空间信息, 并补偿了后续两个级联内核大小和为 $\left\lfloor \frac{k}{d} \right\rfloor \times \left\lfloor \frac{k}{d} \right\rfloor$

的深度扩展卷积。 $\lfloor \cdot \rfloor$ 表示地板操作, 深度扩展卷积负责捕获深度卷积输出  $\bar{Z}^C$  的全局空间信息得到  $Z^C$ 。 $A^C$  注意力图表示将深度卷积的输出进行  $1 \times 1$  核卷积, 最终 LSKA 的输出  $\bar{F}^C$  由注意力图  $A^C$  与输入特征图  $F^C$  进行 Hadamard 积得到。

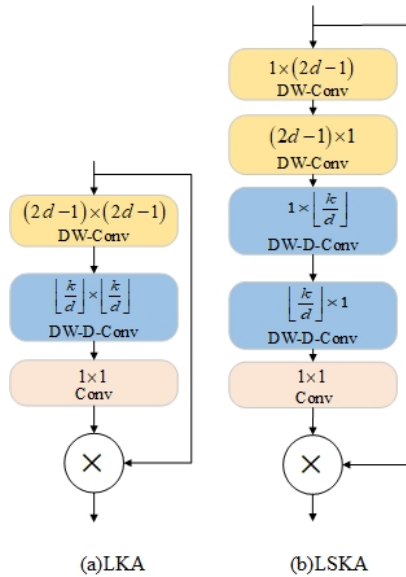


图5 注意力机制对比

Fig.5 Comparison of attention mechanisms

原 LKA 的参数和 FLOPs 计算方法如下:

$$\text{Param} = (2d-1)^2 \times C + \left\lfloor \frac{k}{d} \right\rfloor^2 \times C + C \times C \quad (7)$$

$$\text{FLOPs} = \left( (2d-1)^2 \times C + \left\lfloor \frac{k}{d} \right\rfloor^2 \times C + C \times C \right) \times H \times W \quad (8)$$

式中:  $k$  是核大小;  $d$  是膨胀率。LSKA 模块的 FLOPs 总数和参数计算方法如下:

$$\text{Param} = (2d-1) \times C \times 2 + \left\lfloor \frac{k}{d} \right\rfloor \times C \times 2 + C \times C \quad (9)$$

$$\text{FLOPs} = \left( (2d-1) \times C \times 2 + \left\lfloor \frac{k}{d} \right\rfloor \times C \times 2 + C \times C \right) \times H \times W \quad (10)$$

通过计算公式可以看出 LSKA 比原始 LKA 设计的深度卷积层中节省一半的  $2d-1$  个参数。同样, LSKA 比原始 LKA 设计的深度扩展卷积层中节省一半的  $\left\lfloor \frac{k}{d} \right\rfloor$  个参数。在 FLOPs 数量方面的节省与参数相同。

本文在检测大目标的颈部加入大可分离核注意力机制 (LSKA), 通过捕获图像中的远程依赖关系使模型能够更精确地关注重要信息, 提高了模型对于复杂数据的处理能力。LSKA 注意力机制模块加入改进网络的位置如图 1 所示。

## 2 实验结果与分析

### 2.1 实验环境与数据

本次实验在 GPU 为 Nvidia Geforce RTX 3090, 显存为 24 GB 的服务器上进行, 采用 Pytorch 框架进行测试与实验, 环境配置为 Python3.9+Pytorch 1.7.0+CUDA 11.0。训练 300 批次, 使用随机梯度下降 SGD (Stochastic Gradient Descent) 优化器进行训练, batchsize 设为 16, 初始学习率为 0.01。

本次实验使用的数据集为自建数据集, 图像数据来源于网络爬虫, 通过 Labellmg 对图像进行标签标注。数据集一共有 5455 张图像, 数据集分为 4 类, 分别是海星、海参、海胆和扇贝。数据集样图如图 6 所示。将数据集按 8:1:1 的比例分成训练集、测试集和验证集, 从数据集中随机选取 4365 张图像作为训练集, 545 张图像作为测试集, 545 张图像作为验证集。



图6 数据集样本

Fig.6 Dataset sample

### 2.2 评价指标

用精确度 (Precision)、召回率 (recall) 和平均精确度 (mAP) 来衡量检测的准确性。帧率 (FPS) 用于测量模型的速度, 以时间的倒数计算。Giga 浮点

运算 (GFLOPs), 即 10 亿次浮点运算, 被用作衡量 GPU 性能的指标。各指标的计算公式如下所示:

$$P = \frac{TP}{TP+FP} \tag{11}$$

$$R = \frac{TP}{TP+FN} \tag{12}$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \tag{13}$$

$$AP = \int_0^1 P(R) dR \tag{14}$$

$$FPS = \text{Number/time} \tag{15}$$

式中: TP 是正样本被正确预测为正类的数量; FP 是负样本被错误预测成正类的数量; FN 为原本为正样本但被检测为负类的数量; mAP 为平均精度均值; n 为检测的类别数量; AP 为  $P(R)$  曲线下的面积。

2.3 注意力机制对比实验

为了更好地验证本次改进加入的 LSKA 注意力机制对改进网络性能的影响, 设计了以 YOLOv8n 为基线, 将基线网络分别加入 SE、MLCA、EMA、simAM 注意力机制的对比试验, 对比实验结果如表 1 所示。

表 1 不同注意力机制对比实验

Table 1 Comparison experiment of different attention mechanisms

Algorithms	mAP50/%	mAP50-95/%	Params/MB	GFLOPs/G
YOLOv8n	81.7	47.4	3.0	8.2
YOLOv8n +SE	82.7	48.9	3.0	8.2
YOLOv8n +MLCA	82.5	48.3	3.0	8.2
YOLOv8n +EMA	82.4	48.7	3.0	8.2
YOLOv8n +simAM	82.1	48.4	3.0	8.2
YOLOv8n +LSKA	83.3	49.2	3.1	8.2

从表 1 中可以看出, 与 YOLOv8n 基线相比, 所有加入了注意力机制的网络 mAP 指标都上升了, 且计算复杂度都保持与基线相同。相比于加入其他注意力机制, 加入 LSKA 注意力机制的 YOLOv8 网络检测精度提高明显, 比加入 SE、MLCA、EMA 和 simAM 注意力机制的网络 mAP50 检测精度分别提升 0.6%、0.8%、0.9%和 1.2%, mAP50-95 检测精度分别提升了 0.3%、0.9%、0.5%和 0.8%。虽然加入 LSKA 注意力机制的网络参数量略微上涨, 但是网络整体精度上涨显著。因此考虑到网络的检测精度与检测速度的平衡

性, 加入 LSKA 注意力机制的改进方法使 YOLOv8 网络模型更适用于水下目标检测场景

2.4 AADS 模块在不同位置的对比实验

为了验证 AADS 轻量级下采样模块在网络中的不同位置产生的不同网络影响, 设计了如表 2 所示的消融实验。

AADS-all 表示将主干网络和颈部网络中的卷积模块替换为 AADS 下采样模块; AADS-backbone 表示将主干网络中的卷积模块替换为 AADS 下采样模块, 颈部网络卷积模块部分不替换; AADS-neck 表示将颈部网络中的卷积模块替换为 AADS 下采样模块, 主干网络中的卷积模块部分不替换。根据实验结果来看, 将颈部网络中卷积模块替换为 AADS 下采样模块, 平均精度提高不显著且计算复杂度下降也不明显。将主干网络以及颈部网络的卷积模块替换为 AADS 下采样模块和将主干网络中的卷积模块替换为 AADS 下采样模块这两个实验中, mAP50 平均精度都显著提高, 且参数量和计算复杂度相同, 相比原网络来说参数量和计算复杂度都得到进一步降低。但是将主干网络中的卷积模块替换为 AADS 下采样模块使网络中的 mAP50-95 平均精度提升更多, 因此, 将主干网络中的卷积模块替换为 AADS 下采样模块能使网络获得更好的性能。

表 2 AADS 不同位置对比结果

Table 2 Comparison results of AADS at different positions

Algorithms	mAP50/%	mAP50-95/%	Params/MB	GFLOPs/G
YOLOv8n	81.7	47.4	3.0	8.2
AADS-all	82.7	48.5	2.7	7.9
AADS-backbone	82.7	48.9	2.7	7.9
AADS-neck	82.2	48.4	3.0	8.1

2.5 消融实验

验证改进方法的有效性是非常重要的, 因此在数据集上进行了消融实验。表 3 为改进算法在同一实验环境下的消融实验结果。

由表 3 实验结果可知, 实验 1 为 YOLOv8n 基线模型实验结果, 实验 2 将 BiFPN 双向特征金字塔结构加入到 YOLOv8n 的颈部中, 可以看出 mAP50 平均精度比基线模型提升了 1.3%, 参数量减少了 33.3%, GFLOPs 从 8.2 下降到 7.1, 各方面提升效果显著。实验 3 将 YOLOv8n 主干网络中的卷积模块替换为轻量级 AADS 下采样模块, 从实验结果上可以看出, mAP50 平均精度提升了 1%, 参数量和计算复杂度都明显降低, 进一步减少网络的参数量和计算复杂度。实验 4 在 YOLOv8n 基线模型预测大尺寸目标的颈部

网络与头部网络之间加入 LSKA 注意力机制，虽然参数量略微上升，但是在同等网络计算复杂度的情况下，mAP50 平均精度从 81.7%提升到 83.3%，提升了 1.6%，精度提升效果最好，且使模型参数量上涨不高，证明了在 YOLOv8n 网络中加入 LSKA 注意力机制是有效的。实验 5、6、7 分别为将 3 个改进点随机两两结合得到的实验结果，由实验结果可知，3 个改进实验均使 YOLOv8 基线模型的平均检测精度和 FPS 提升，模型参数量和计算量都有一定程度的下降，但是

没有使各指标达到最优的平衡效果。实验 8 为在基线模型基础上加入所有改进方法，mAP50 平均精度虽然没有达到最优，但是相比于基线模型的平均精度提升了 1.4%，达到了次优，精度提升明显，参数量和 GFLOPs 都降到最低，相比基线模型分别降低了 43.3%和 15.9%，FPS 检测速度从 108f/s 提升至 114f/s。总的来说，提出的改进方法有效地提升了 YOLOv8 网络性能，在水下目标检测中实现了检测精度与速度的平衡。

表 3 消融实验结果

Table 3 Results of ablation experiment

Models	BiFPN	AADS	LSKA	mAP50/%	mAP50-95/%	Params/MB	GFLOPs/G	FPS/(f/s)
1				81.7	47.4	3.0	8.2	108
2	√			83.0	49.0	2.0	7.1	112
3		√		82.7	48.9	2.7	7.9	110
4			√	83.3	49.2	3.1	8.2	104
5	√	√		82.8	48.1	1.7	6.9	117
6	√		√	82.9	48.7	2.0	7.1	112
7		√	√	82.3	48.6	2.7	8.0	109
8	√	√	√	83.1	48.8	1.7	6.9	114

2.6 对比实验

为了验证改进算法相比于其他算法的优越性，在同一数据集的情况下对改进算法和当前主流目标检测算法进行了对比实验，实验结果如表 4 所示。

表 4 不同算法对比结果

Table 4 Comparison results of different algorithms

Algorithms	mAP50/ %	Params/ MB	GFLOPs/ G	FPS/ (f/s)
SSD	74.4	26.3	62.8	26
Faster-RCNN	71.2	137	370	12
RetinaNet	58.4	38	170.1	23
YOLOv3	73.0	61.9	66.2	25
YOLOv4-tiny	63.6	6.1	7.0	56
YOLOv5s	82.7	7.1	16.1	63
Ref. [28]	80.9	4.5	8.5	68
YOLOv7-tiny	81.6	6.0	13.2	70
Ours	83.1	1.7	6.9	114

其中对比的单阶段目标检测算法包括 SSD，RetinaNet 和 YOLO 系列算法；两阶段目标检测算法为 Faster-RCNN。

通过表 4 的对比实验可以看出，改进的目标检测算法在各性能方面都优于其他主流目标检测算法。与单阶段目标检测检测算法 SSD 和 RetinaNet 相比，改进的算法在各个性能表现上都提高明显，尤其对于

RetinaNet 网络来说，mAP50 平均精度提高了 24.7%，计算复杂度下降了 163.2 G。与传统两阶段目标检测算法 Faster-RCNN 相比，改进的算法检测精度和检测速度方面都提升显著，参数量和计算复杂度更是下降明显。与 YOLOv3 网络相比，参数量降低了近 36 倍，计算复杂度降低了近 10 倍，mAP50 提升了 10.1%。与轻量级 YOLOv4-tiny 和 YOLOv7-tiny 网络相比，mAP50 分别提升 19.5%和 1.5%，参数量都降低约 72%，计算复杂度分别降低了 0.1 G 和 6.3 G，推理速度分别提升 58 f/s 和 44 f/s。对于 YOLOv5s 网络来说，虽然 mAP50 平均精度提升不明显，但是参数量降低了 76%，计算复杂度下降了 57%，推理速度提高了 51 f/s。与文献[28]中的算法相比，mAP50 平均精度提升 2.2%，模型参数量和计算复杂度分别降低 62.2%和 18.8%，FPS 提高近两倍，改进网络在的各个方面都占优。

2.7 改进算法检测效果实验

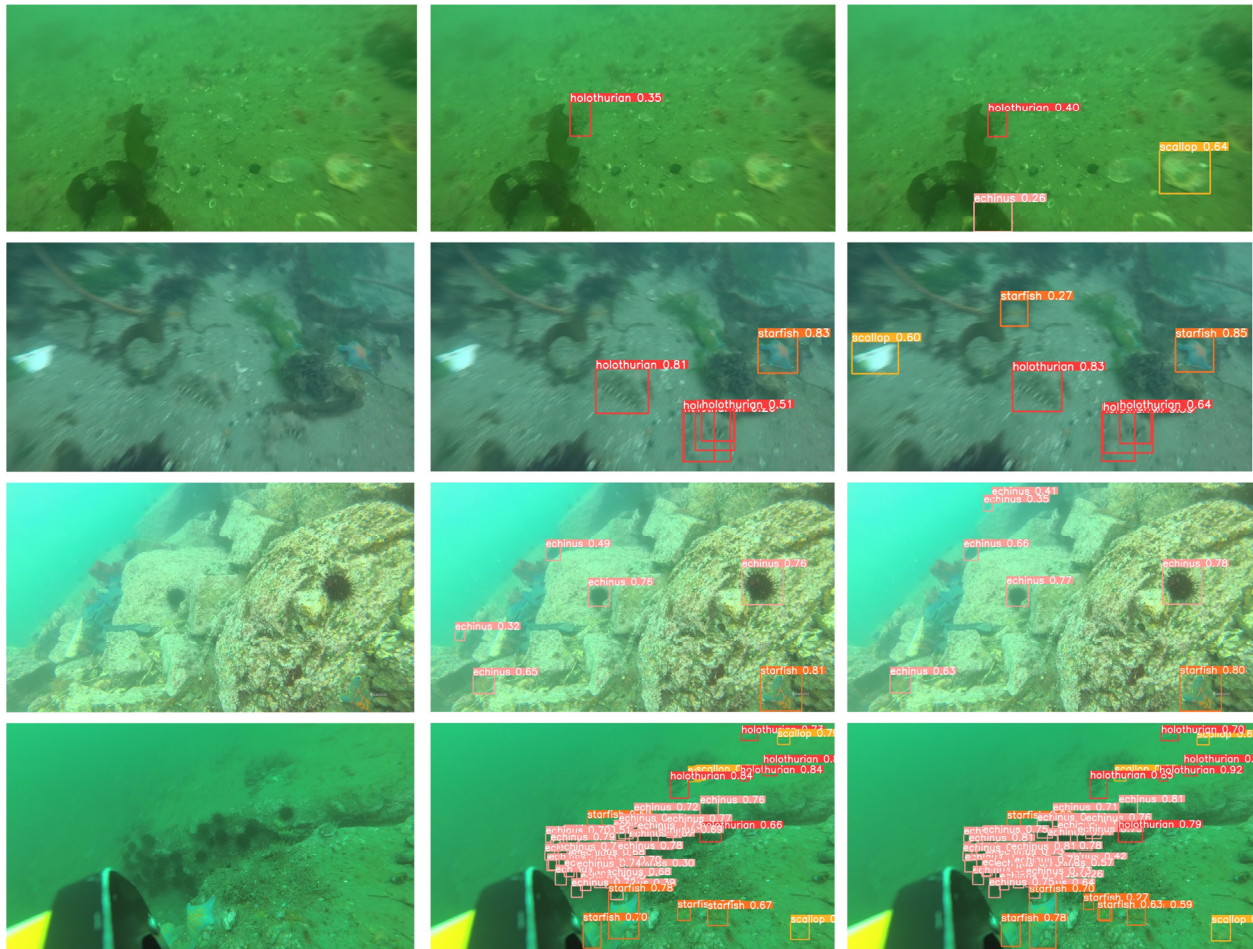
为了进一步直观地展现改进算法的检测效果，使用改进的网络和 YOLOv8n 网络分别在数据集上进行检测实验，实验效果图如图 7 所示。

图 7 第一列图片为水下数据集中的原图照片，第二列图片为在 YOLOv8n 基线模型上的检测效果图，第三列图片为在改进算法上的检测效果图。从图 7 效果对比图中可以看出 YOLOv8n 网络模型存在误检和



漏检行为，并且检测精度不高。从第一行和第二行检测对比图片中可以看出，原基线模型存在漏检行为，改进的算法不仅提高了检测精度，并且减少了漏检行为。从第三行对比图片中，YOLOv8n 模型误检出左下角海胆信息，改进的算法不仅消除误检行为并且检测出漏检的远处小目标海胆信息。第四行效果对比图

中,改进算法相对于 YOLOv8n 基线模型对密集小目标预测框的选择范围更精确,且正确预测率较高。综上所述,改进的算法网络不仅使检测精度提高,而且原模型中漏检的目标也能检测出来,在重叠目标和小目标的图片检测中表现更好,大大提高了检测性能。因此,改进的网络算法能更好地检测水下目标。



(a) 原图 (b) YOLOv8n 检测图 (c) 改进算法检测图  
(a) Original graph (b) YOLOv8n detection graph (c) Improve algorithm detection graph

图 7 检测效果对比

Fig.7 Comparison of detection effects

### 3 结论

针对水下图像检测效率低、应用于移动设备困难等问题,本文提出了一种改进 YOLOv8 网络的轻量化水下目标检测算法。在颈部网络融合 BiFPN 特征金字塔结构提高小目标检测精度,设计轻量化下采样模块 AADS 使整体网络轻量化,并且引入 LSKA 注意力机制捕获图像特征中的重要信息。将改进算法在数据集上进行消融实验,改进后的模型与原模型相比,参数量下降 43.3%,计算复杂度降低 15.9%,并且 mAP50

比原模型提高了 1.4%，FPS 达到 114 f/s，实现了网络的轻量化，检测精度与检测速度之间达到了平衡。并与其他算法进行了对比实验，结果表明改进的算法平均检测精度更高，模型参数量与计算复杂度下降明显，在重叠目标和小目标的图片检测中效果突出，整体检测性能进一步提高，更加适合复杂的水下环境。下一步工作中将继续优化算法，进一步提高算法的鲁棒性和准确性，将改进算法部署到移动设备上。探索更多轻量化检测方法在水下目标检测方面的应用。

## 参考文献:

- [1] SHEN Linghao, XIA Haisheng, ZHANG Xun, et al. U2PNet: an unsupervised underwater image-restoration network using polarization[J]. *IEEE Transactions on Cybernetics*, 2024: 1-14. Doi: 10.1109/TCYB.2024.3365693.
- [2] Yeh C H, LIN C H, KANG L W, et al. Lightweight deep neural network for joint learning of underwater object detection and color conversion[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **33**(11): 6129-6143.
- [3] 周辉奎, 章立, 胡素娟. 改进直方图匹配和自适应均衡的水下图像增强[J]. *红外技术*, 2024, **46**(5): 532-538.
- ZHOU Huikui, ZHANG Li, HU Sujuan. Underwater image enhancement based on improved histogram matching and adaptive equalization[J]. *Infrared Technology*, 2024, **46**(5): 532-538.
- [4] SHEN L, Reda M, ZHANG X, et al. Polarization-driven solution for mitigating scattering and uneven illumination in underwater imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, **62**: 4202615.
- [5] XU S, ZHANG M, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection[J]. *Neurocomputing*, 2023, **527**: 204-232.
- [6] REN S, HE K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **39**(6): 1137-1149.
- [7] LIU W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//*Computer Vision—ECCV*, 2016: 21-37.
- [8] 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[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464-7475.
- [9] ZENG L, SUN B, ZHU D. Underwater target detection based on Faster R-CNN and adversarial occlusion network[J]. *Engineering Applications of Artificial Intelligence*, 2021, **100**: 104190.
- [10] SONG P, LI P, DAI L, et al. Boosting R-CNN: reweighting R-CNN samples by RPN's error for underwater object detection[J]. *Neurocomputing*, 2023, **530**: 150-164.
- [11] YU G, CAI R, SU J, et al. U-YOLOv7: a network for underwater organism detection[J]. *Ecological Informatics*, 2023, **75**: 102108.
- [12] HUA X, CUI X, XU X, et al. Underwater object detection algorithm based on feature enhancement and progressive dynamic aggregation strategy[J]. *Pattern Recognition*, 2023, **139**: 109511.
- [13] XU X, LIU Y, LYU L, et al. MAD-YOLO: a quantitative detection algorithm for dense small-scale marine benthos[J]. *Ecological Informatics*, 2023, **75**: 102022.
- [14] Terven J, Cordova-Esparza D. A comprehensive review of YOLO: from YOLOv1 to YOLOv8 and beyond[J]. arxiv preprint arxiv:2304.00501, 2023.
- [15] Al Muksit A, Hasan F, Emon M F H B, et al. YOLO-Fish: a robust fish detection model to detect fish in realistic underwater environment[J]. *Ecological Informatics*, 2022, **72**: 101847.
- [16] CHEN L, ZHENG M, DUAN S, et al. Underwater target recognition based on improved YOLOv4 neural network[J]. *Electronics*, 2021, **10**(14): 1634.
- [17] LIU P, QIAN W, WANG Y. YWnet: A convolutional block attention-based fusion deep learning method for complex underwater small target detection[J]. *Ecological Informatics*, 2024, **79**: 102401.
- [18] YU G, CAI R, SU J, et al. U-YOLOv7: a network for underwater organism detection[J]. *Ecological Informatics*, 2023, **75**: 102108.
- [19] Fayaz S, Parah S A, Qureshi G J, et al. Intelligent underwater object detection and image restoration for autonomous underwater vehicles[J]. *IEEE Transactions on Vehicular Technology*, 2024, **73**(2): 1726-1735.
- [20] Talaat F M, Zain Eldin H. An improved fire detection approach based on YOLO-v8 for smart cities[J]. *Neural Computing and Applications*, 2023, **35**(28): 20939-20954.
- [21] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, **34**(7): 12993-13000.
- [22] LI X, WANG W, WU L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. *Advances in Neural Information Processing Systems*, 2020, **33**: 21002-21012.
- [23] LIN T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2117-2125.
- [24] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8759-8768.
- [25] ZHANG X, LIU C, YANG D, et al. Rfaconv: Innovating spatital attention and standard convolutional operation[J]. arxiv preprint arxiv:2304.03198, 2023.
- [26] Ioannou Y, Robertson D, Cipolla R, et al. Deep roots: improving CNN efficiency with hierarchical filter groups[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1231-1240.
- [27] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[J]. *Computational Visual Media*, 2023, **9**(4): 733-752.
- [28] SUN Y, ZHENG W, DU X, et al. Underwater small target detection based on YOLOX combined with MobileViT and double coordinate attention[J]. *Journal of Marine Science and Engineering*, 2023, **11**(6): 1178.