

改进 YOLOv11 的轻量化水下生物目标检测算法*

郭文豪 郝 斌 张 飞

(内蒙古科技大学数智产业学院 包头 014010)

摘 要:水下生物目标检测技术对于海洋生态保护、渔业资源开发等众多领域的发展具有重要意义。为了提高水下生物检测精度和速度以及解决现有水下生物检测模型参数量较大的问题,提出了一种改进 YOLOv11 的轻量化水下生物检测算法 YOLOv11-PiA。首先,通过引入部分卷积(partial convolution, PConv)对 C3k2 模块进行改进,设计出 C3k2_PConv 模块,有效降低了模型的参数量和计算复杂度,并提高了检测精度和速度。其次,结合 C2PSA 模块和倒置残差块注意力机制(inverted residual block attention mechanism, iRMB),提出了 C2PSA_iRMB 模块,该模块通过强化不同深度的特征信息以增强特征表达能力,进而提升了检测精度。最后,采用轻量化下采样模块(lightweight downsampling module, ADown)替换基线模型中的传统卷积下采样模块,模型的参数量得到进一步降低,检测效率进一步提升。实验数据表明,在水下生物数据集 RUOD 和 DUO 上, YOLOv11-PiA 模型的 mAP@0.5 达到了 85.6% 和 82.6%, 帧率(frames per second, FPS) 达到了 97 和 116 fps。与 YOLOv11n 原模型相比, YOLOv11-PiA 模型的 mAP@0.5 分别提高了 2.8% 和 2.3%, FPS 分别提高了 9 和 13 fps, 且参数量、浮点运算量和模型大小分别降低了 29.8%、22.2% 和 29.1%。其综合性能优于 RTDETR、YOLOv12 等主流的目标检测模型, 可实现对水下生物的快速准确检测。

关键词:水下生物目标检测; YOLOv11; 轻量化; PConv; 注意力机制

中图分类号: TN911.73; TP391

文献标识码: A

国家标准学科分类代码: 520.6040

Improved YOLOv11 lightweight underwater biological object detection algorithm

Guo Wenhao Hao Bin Zhang Fei

(School of Digital and Intelligent Industry, Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: Underwater biological target detection technology is of great significance to the development of numerous fields, such as marine ecological protection and fishery resource exploitation. To improve the accuracy and speed of underwater biological detection and address the high parameter count in existing models, a lightweight algorithm, YOLOv11-PiA, based on improved YOLOv11, is proposed. Firstly, the C3k2 module is improved by introducing partial convolution (PConv), resulting in the C3k2_PConv module, which effectively reduces the model's parameter count and computational complexity while enhancing detection accuracy and speed. Secondly, the C2PSA_iRMB module, which combines the C2PSA module with the inverted residual block attention mechanism (iRMB), is proposed. This module strengthens feature information at multiple depths to enhance feature representation, thereby improving detection accuracy. Finally, the traditional convolutional downsampling module in the baseline model is replaced with the lightweight downsampling module (ADown), further reducing the model's parameter count and improving detection efficiency. Experimental results on the RUOD and DUO underwater biological datasets show that the proposed YOLOv11-PiA model achieves mAP@0.5 values of 85.6% and 82.6%, with frames per second (FPS) reaching 97 and 116 fps, respectively. Compared with the original YOLOv11n model, YOLOv11-PiA increases mAP@0.5 by 2.8% and 2.3%, and improves FPS by 9 and 13 fps, respectively. Meanwhile, the parameter count, floating-point operations per second (FLOPs), and model size are reduced by 29.8%, 22.2%, and 29.1%, respectively. The comprehensive performance of YOLOv11-PiA outperforms mainstream target detection models such as RTDETR and

YOLOv12, enabling fast and accurate detection of underwater organisms.

Keywords: underwater biological target detection; YOLOv11; lightweight; PConv; attention mechanism

0 引言

建设海洋强国是国家长期发展战略,而海洋观测仪器设备不仅是建设海洋强国的重要抓手,也是国家整体海洋技术水平的体现^[1]。近年来,随着目标检测技术的飞速进步,水下生物目标检测技术在海洋观测仪器设备中的应用愈发广泛。从海洋资源的可持续勘探与开发利用,到海底生态系统中关键生物种群的健康监测与保护;从辅助海洋环境保护与生态修复项目的精准实施,到渔业资源的科学管理与可持续捕捞策略的制定,水下生物目标检测技术均扮演着不可或缺的角色,这一技术的应用不仅能够帮助科研人员更深入地理解海洋生物多样性的分布与变迁,为生态保护提供科学依据,还可有效提升海洋资源开发的效率和安全性,促进海洋经济的健康发展。在渔业领域,通过精准识别鱼类及其他海洋生物的种类、数量及分布,该技术有助于实现渔业资源的合理调配与保护,避免过度捕捞,维护海洋生态平衡。此外,在海洋环境监测与灾害预警方面,水下生物目标检测也能发挥重要作用,比如通过监测珊瑚礁生态系统的健康状况,预警海洋酸化、水温异常等环境变化对生物群落的影响,为制定应对措施赢得宝贵时间。因此,水下生物目标检测技术不仅具备突出的科研价值,更是助力海洋经济实现可持续发展、维护海洋生态系统安全的核心技术,对该技术开展系统性的深入研究与推广应用,具有极为广阔的发展前景,值得科研人员持续探索与投入。

深度学习技术的问世促进了计算机视觉领域的发展。其中,以区域卷积神经网络(region-based convolutional neural network, R-CNN)^[2]为典型代表的两阶段目标检测算法,以及以单发多框检测器(single shot multibox detector, SSD)^[3]、YOLO(you only look once)系列模型^[4]为核心的单阶段目标检测算法,为水下生物目标检测任务开辟了全新的研究维度,诸多科研工作者以此为依托,开展了一系列系统性的探索与改进工作。Shi等^[5]针对水下生物目标检测的难题,提出了基于 Faster-RCNN 的改进算法。该算法采用残差网络(residual network, ResNet)作为骨干特征提取网络;通过双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)构建 ResNet-BiFPN 结构以增强特征提取与多尺度融合能力;用增强交并比(enhanced intersection over union, EIou)替代交并比(intersection over union, IoU)减少训练数据中冗余边界框;利用 K-means++ 聚类生成更适配的锚框,这些改进有效提升了模型的检测精度并降低了漏

检率。但两阶段架构固有计算复杂度高,难以适配水下检测设备的资源约束;且 EIou 对水下低对比度目标的边界框优化效果有限,模型的漏检率在小目标场景下仍然较高。Huang 等^[6]针对水下图像质量不佳所引发的水下生物检测精度偏低这一关键问题,提出了一种基于改进 SSD 的水下生物目标检测算法 MFFNet。该算法通过对 VGG-16 和 ResNet-50 进行复合主干连接,引入卷积块注意力模块(convolutional block attention module, CBAM)以及利用特征金字塔网络(feature pyramid network, FPN)结构进行多尺度特征检测,有效提升了模型在水下生物检测任务中的表现。但 VGG-16 骨干网络的冗余参数并未得到精简,模型参数量较大且推理速度较慢,无法满足实时检测需求;同时 FPN 对水下多尺度目标的特征适配性不足,跨尺度检测精度波动较大。史朋飞等^[7]提出了一种结合数据增强和改进 YOLOv4 的水下生物目标检测算法。该算法在主干网络 CSPDarknet53 中添加了 CBAM 注意力机制,并在路径聚合网络(path aggregation network, PANet)中添加了同层跳接和跨层跳接结构,通过数据增强方法模拟了水下生物重叠、遮挡等显示不完全的情形,从而实现了检测精度的大幅提升。但 CSPDarknet53 骨干网络未做轻量化改进,模型在嵌入式设备上部署困难;且数据增强仅模拟了重叠遮挡场景,未针对水下浑浊、光线不均导致的特征模糊问题进行优化,模型在复杂环境下的误检率偏高。张银胜等^[8]则将 YOLOv5s 算法应用于水下场景并进行针对性改进。首先借助 Double MSRCR 方法改善水下物体清晰度不足、特征模糊的问题,其次在 YOLOv5s 中融入 C_VAN 模块、RFB_S 感受野以及归一化注意力模块(normalization attention module, NAMAttention),以此增强模型的多尺度特征提取能力。实验证明,该算法在水下生物目标检测中的表现优于常规检测模型,适用于复杂水下环境。但 Double MSRCR 算法的计算开销较大,这导致模型的推理速度下降;且未针对水下小目标的特征增强做专项优化,小目标检测精度较低。针对复杂弱光环境下水下生物资源识别能力低、特征丢失严重等问题,周新等^[9]提出了轻量级水下生物检测算法 YOLOv5_PGS。该算法通过暗通道-对比限制-光衰减算法以丰富图像特征信息;引入 GhostNet 网络架构,并针对性构建 C3CA 特征模块,以此提升模型的特征提取效率与多维度特征融合性能;优化损失函数以降低总损失自由度、增强模型泛化能力,这些改进实现了模型在水下生物检测中精度与速度的平衡,为水下生物的实时检测提供了可靠保障。但 GhostNet 的特征映射存在信息损失,模型对水下生物的细节特征

表达不足,类别混淆率较高;且暗通道增强算法易放大水下噪声,这会导致模型在复杂场景下的鲁棒性不足。Yi 等^[10]针对水下生物目标检测中漏检率高、场景识别能力差的问题,提出了一种名为 UWSC-YOLOv7 的水下生物目标检测算法。该算法以 YOLOv7 为基础网络,通过融合挤压激励网络(squeeze-and-excitation network, SENet)、增强 FPN 网络拓扑结构以及结合 EIoU 损失函数,在兼顾检测速度的同时,成功提升了检测精度,为水下生物检测提供了高效的解决方案。但 SENet 注意力机制仅聚焦通道维度,未考虑空间维度的特征关联,对水下目标的局部细节特征捕捉不足,这导致模型的漏检率较高;且 YOLOv7 的骨干网络未做轻量化裁剪,模型参数量较大,难以适配资源受限的水下检测设备。针对水下生物检测仪器存储和计算资源有限、高参数量模型占用资源多的问题,苗力恒等^[11]提出了一种改进 YOLOv8n 的轻量化水下生物目标检测算法 YOLOv8n-MAL。该算法首先通过多尺度卷积块注意力模块(multi-scale convolutional block attention module, MSCBAM)提升模型在复杂场景中的鲁棒性,其次设计新型 MSFFN 结构和 PC2F-LMS 模块提升模型的多尺度特征提取能力,最后采用加权交并比(weighted intersection over union, WIoU)优化原模型的完全交并比(complete intersection over union, CIoU)。实验证明,该算法检测精度高且实现了轻量化。但 MSCBAM 注意力机制的计算复杂度偏高,这导致模型的推理速度较低;此外, WIoU 损失函数对低 IoU 阈值的目标优化不足,模型易出现漏检。Pan 等^[12]针对水下图像模糊、环境复杂导致的特征提取难与漏检问题,提出了一种改进 YOLOv10 的轻量级水下生物目标检测算法 YOLOv10-AD。该算法通过引入专为水下生物形状优化的 AKVVanillaNet 主干网络,在提升检测精度的同时大幅减少了参数与计算成本,并在 C2f 结构中集成动态蛇形卷积(dynamic snake convolution, DysnakeConv)以增强特征提取能力,最后采用了预测交并比(predicted intersection over union, PIoU)优化数据拟合效果。实验结果表明,该算法适用于水下复杂环境,具有检测精度高且速度快的特点。但 PIoU 损失函数未结合水下目标的尺度差异,这会导致大目标与小目标的检测精度失衡。

综上,现有水下生物目标检测算法普遍存在 3 大核心局限:轻量化与高精度难以平衡(要么参数量大、要么小目标检测精度不足);特征提取模块对水下复杂环境(噪声、低对比度、多尺度)的适配性不足;下采样过程中细节特征丢失严重,影响边界框回归精度。鉴于 YOLOv11^[13]兼具高精度与高效率的优异性能,本文提出了一种改进 YOLOv11 的轻量化水下生物目标检测算法 YOLOv11-PiA,旨在提高水下生物目标检测的准确性和稳定性。首先,针对传统卷积计算冗余的问题,在 C3k2

模块中引入高效的部分卷积(partial convolution, PConv)^[14],设计 C3k2_PConv 模块。该模块通过选择性卷积策略,可在维持模型优异特征提取性能的同时降低模型的参数量与计算复杂度,同时实现检测精度与推理速度的双重提升。其次,为解决水下场景特征表达能力不足的问题,将 C2PSA 模块与倒置残差块注意力机制(inverted residual block attention mechanism, iRMB)^[15]融合,提出了具有多尺度特征感知能力的 C2PSA_iRMB 模块。该模块通过强化深层次特征的交互增强模型对水下生物多尺度特征的表达能力,进而提升检测精度。最后,针对下采样过程中的特征信息丢失问题,采用轻量化下采样模块(lightweight downsampling module, ADown)^[16]替代传统卷积下采样模块。该模块通过高效的双分支特征处理策略,不仅可维持特征图的信息完整性,还能有效降低下采样过程中的计算开销。

1 YOLOv11 模型介绍

2024 年 9 月, YOLOv11 模型正式发布,该模型依据参数量与计算复杂度的差异,衍生出 5 个版本。本研究选取了在 COCO 数据集^[17]上兼具高精度检测性能与轻量化特性的 YOLOv11n 作为改进的基线模型。该模型的网络架构主要包含输入层(input)、骨干特征提取网络层(backbone)、颈部特征融合层(neck)以及头部检测层(head)4 大核心模块,其具体结构如图 1 所示。

输入层作为整个模型的起点,主要是将数据集的图片进行一系列数据预处理操作,以便为后续网络层的特征提取和模式识别提供规范化的输入。骨干网络层集成了 Conv、C3k2、SPPF、C2PSA 模块,其核心功能是通过卷积操作与池化操作完成对输入图片特征的提取与整合。颈部层集成了 Concat、Upsample 等模块,通过这些模块对骨干网络提取的多尺度特征进行进一步融合和优化,随后将其传递至头部层以执行预测任务。头部层部分通过卷积运算完成特征映射,将目标特征信息转化为检测任务所需的类别置信度分数与边界框坐标回归偏移量。

2 YOLOv11-PiA 模型介绍

2.1 YOLOv11-PiA 模型概述

针对水下生物目标检测面临的问题,本文提出了一种适用于水下生物检测的 YOLOv11-PiA 模型,其结构如图 2 所示。主要改进为:首先引入部分卷积 PConv 对 C3k2 模块进行改进,设计 C3k2_PConv 模块,在有效降低模型的参数量和计算复杂度的同时,提高检测精度与速度;其次,将 C2PSA 模块和倒置残差块注意力机制 iRMB 结合,提出 C2PSA_iRMB 模块,该模块通过强化不同深

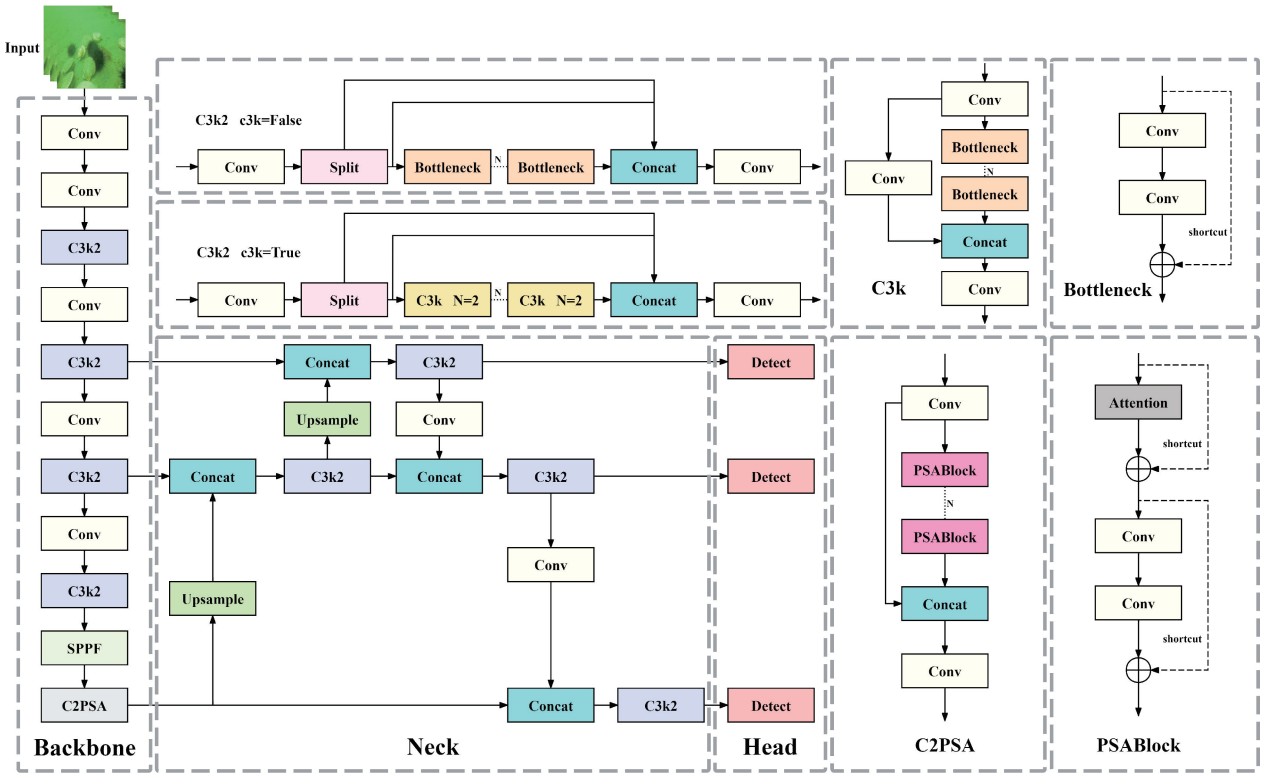


图 1 YOLOv11n 模型结构

Fig. 1 YOLOv11n model structure diagram

度的特征信息以增强特征表达能力,进而实现检测精度的提升;最后采用轻量化下采样模块 ADown 替换基线模型中的传统卷积下采样模块,进一步降低模型的参数与计算量,提升检测效率。

2.2 C3k2_PConv 模块

YOLOv11 模型主要使用传统卷积和 C3k2 模块对图像进行高质量的特征提取。C3k2 模块作为传统 CSP Bottleneck 结构的优化变体,依托并行卷积分支的创新设计与灵活可调的参数配置方案,具备更为卓越的特征提取性能,尤其适用于特征提取任务层级较深、应用场景复杂度较高的实际需求。在并行卷积分支的设计框架下,输入特征图被拆分为两个分支:其一直接进行特征传递,以此完整保留浅层特征的原始信息;其二则通过堆叠若干支持可变卷积核操作的可配置 Bottleneck 或 C3k 模块,完成深层特征的挖掘,最后 2 个分支的特征通过多尺度拼接的方式实现融合。而在参数调节机制层面,该模块可通过设定差异化的 $c3k$ 参数,灵活切换 Bottleneck 或 C3k 模块的组合形式,进而达成计算效率与模型性能之间的动态均衡。

尽管 C3k2 模块通过并行卷积分支设计和灵活参数设置提升了模型的特征提取能力,但其特征提取过程仍以传统卷积和堆叠模块设计为主。这不仅会导致冗余信

息堆积、增加网络计算量,还可能会因为特征信息处理不足而引发误检或漏检问题,影响模型检测性能。而部分卷积 PConv 可以有效解决上述问题,其结构如图 3 所示。

PConv 的计算量 $FLOPs$ 、内存访问量 M 和卷积率 r 计算公式如下:

$$FLOPs_{PConv} = h \times w \times k^2 \times cp^2 \quad (1)$$

$$M = h \times w \times 2cp + k^2 \times cp^2 \approx h \times w \times 2cp \quad (2)$$

$$r = \frac{cp}{c} \quad (3)$$

式中: h 和 w 分别代表输入特征图的长度与宽度; k 代表卷积核的尺寸; c 代表输入特征图的通道数量; cp 代表参与卷积运算的通道数量。

由图 3 和式 (1) ~ (3) 可知, PConv 仅针对输入特征通道中的特定子集开展空间特征提取操作,其余通道则维持原始状态不做任何改变。这一特性使得当参与卷积的通道比例为 $1/4$ 时, PConv 的计算量仅为传统卷积的 $1/16$,且内存访问量约为传统卷积的 $1/4$ 。因此,与传统卷积相比, PConv 不仅没有损失后续特征的通道信息,而且还能切实降低模型的内存占用率,减轻计算负担。鉴于 PConv 在减少特征冗余和内存访问方面的优势,本研究将其引入 C3k2 模块,设计了 C3k2_PConv 模块,其结构如图 4 所示。

首先,通过引入 PConv 对传统 Bottleneck 模块重新设

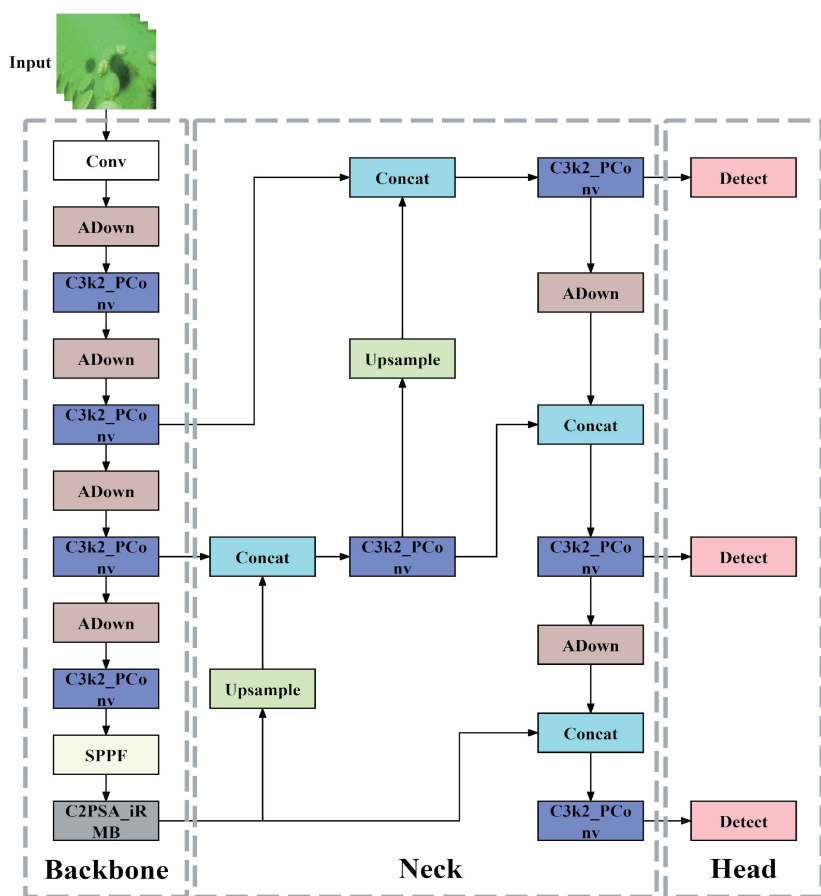


图 2 YOLOv11-PiA 模型结构

Fig. 2 YOLOv11-PiA model structure diagram

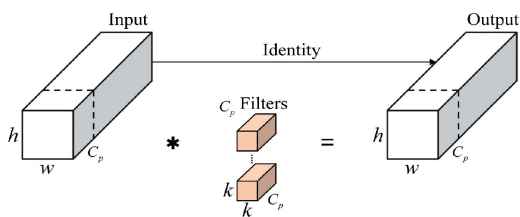


图 3 PConv 模块结构

Fig. 3 PConv module structure diagram

计,得到 P_Bottleneck 模块;其次,将 C3k 模块中的 Bottleneck 模块替换为 P_Bottleneck 模块,构建全新的 C3k_PConv 模块;最后,将 C3k_PConv 模块嵌入 C3k2 模块,设计 C3k2_PConv 模块。通过这一改进,模型在降低参数量与计算复杂度的同时,有效保留了特征提取能力,为水下生物检测任务中精度与速度的平衡提供了保障。

2.3 C2PSA_iRMB 模块

C2PSA 模块是 YOLOv11 骨干网络中新增的核心特征提取组件。在结构设计上,C2PSA 模块沿用了 CSP 结构的分段处理思路,先通过 1×1 卷积将输入特征图划分

为两个分支,其中一个分支直接进行特征传递,另一分支则送入金字塔切片注意力机制(pyramid slice attention, PSA)^[18],最终对两个分支的特征实施拼接融合操作,从而提升模型的特征表示能力。PSA 注意力机制是在 SE 注意力机制^[19]的基础上改进而来的。SE 注意力机制借助对特征通道间关联关系的显式建模,自适应地校正各通道对应的特征响应强度,以此提升模型的特征表征性能。PSA 注意力机制则通过引入金字塔切片的核心设计理念,进一步强化了注意力机制的作用效能,使其可以更精准地捕获特征维度间的复杂依赖关系。此外,PSA 注意力机制采用了 3×3 、 5×5 、 7×7 等多尺度卷积核并行执行特征提取操作,同时融合 SE 注意力机制对通道维度特征实施动态加权处理,强化关键特征通道的响应强度。通过结合 CSP 结构和 PSA 注意力机制,C2PSA 模块有效提升了模型的多尺度特征提取能力,使其在复杂场景下的多尺度目标检测任务中表现出更强的鲁棒性。

为进一步克服性能和复杂性之间的矛盾,并提高模型的学习能力和对目标特征的检测能力,在 C2PSA 模块中引入了 iRMB 注意力机制,使模型更关注水下生物图片中的关键特征,从而提高水下生物目标检测的精度。

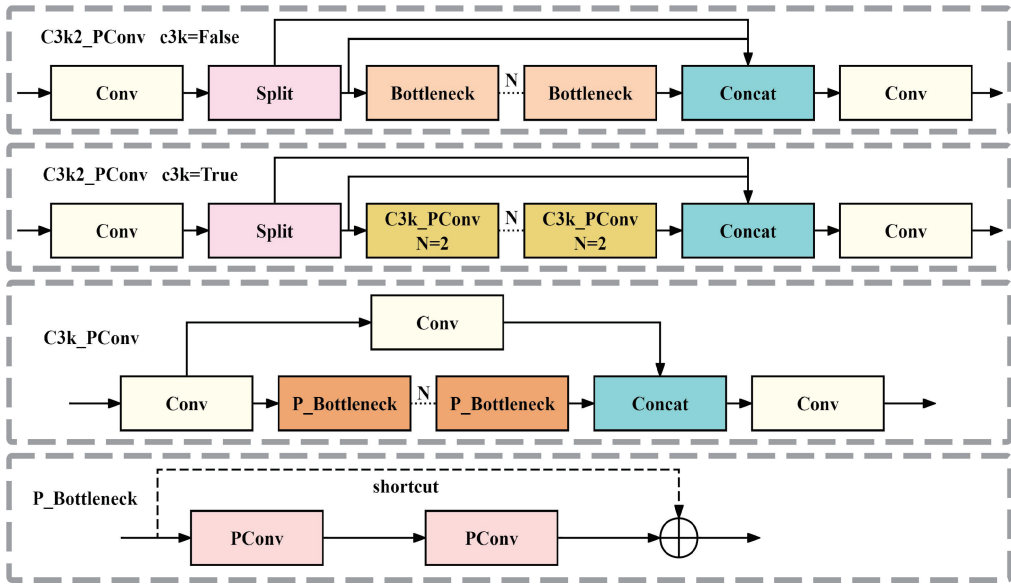


图 4 C3k2_PConv 模块结构

Fig. 4 C3k2_PConv module structure diagram

iRMB 注意力机制借鉴了 MobileNetv2^[20] 中的反向残差结构 (IRB), 同时结合了卷积神经网络在局部特征捕获方面的高效特性, 以及 Transformer^[21] 架构在动态建模长距离特征交互上的独特优势。此外, 该模块借助多头自注意力与深度可分离卷积等核心操作来减少模型的参数和计算量, 在提升模型性能的同时还具备设计上的灵活性, 能够应用于不同的网络架构, 其结构如图 5 所示。

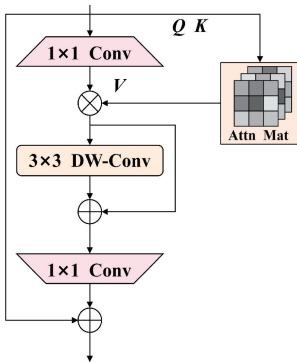


图 5 iRMB 模块结构

Fig. 5 iRMB module structure diagram

由图 5 可知 iRMB 注意力机制的具体流程为: 首先, 通过联合多层感知器构建注意力矩阵 Q 与 K ; 随后, 利用膨胀卷积生成注意力矩阵 V ; 接着, 采用扩展窗口多头自注意力机制对矩阵 Q 、 K 处理, 实现远程信息的有效交互; 之后, 依托深度卷积 (depthwise convolution, DWConv) 完成对局部特征的精准建模; 最后, 利用压缩卷积恢复特征通道维度, 并与初始输入特征构建残差连接, 从而输出

任务所需的最终特征结果。由此可见, iRMB 注意力机制的核心操作是将扩展窗口多头自注意力机制与深度卷积相结合, 具体公式如下:

$$F_{iRMB}(\mathbf{x}) = \text{DW-Conv}(\text{EW-MHSA}(\mathbf{x})) \quad (4)$$

式中: $F_{iRMB}(\mathbf{x})$ 表示 iRMB 注意力机制处理的特征图, $\mathbf{x}$ 表示输入的特征图, $\text{EW-MHSA}(\mathbf{x})$ 表示扩展窗口多头自注意力机制的输出。

在水下生物检测任务中, 场景的复杂性与目标的小尺度特性使得漏检、误检问题频发, 严重影响了检测精度。而 iRMB 注意力机制依托动态全局建模与静态局部信息融合的双重核心优势, 不仅能够精准捕获不同尺度目标的特征信息, 还可有效扩大模型感受野, 从而高效应对水下生物检测场景中存在的各类挑战。鉴于此, 本研究将 iRMB 注意力机制引入 C2PSA 模块, 设计出 C2PSA_iRMB 模块, 其结构如图 6 所示。

由图 6 可知, 本研究首先将 iRMB 注意力机制引入 PSABlock 模块, 构建出 iRPSABlock 模块; 随后将 C2PSA 模块中的所有 PSABlock 模块替换为 iRPSABlock 模块, 从而得到全新的 C2PSA_iRMB 模块。通过此改进, 模型不仅能够复杂多变的水下检测场景中, 精准捕捉细小目标与被遮挡水下生物的核心特征信息, 而且还增强了对水下生物多尺度特征的适应性, 使其在水下生物检测任务中展现出更强的鲁棒性与泛化能力。

2.4 ADown 下采样模块

在目标检测模型中, 下采样是一种通过缩减特征图空间维度, 助力模型捕获更高层级语义特征的关键技术手段。YOLOv11 模型采用传统卷积操作完成下采样过

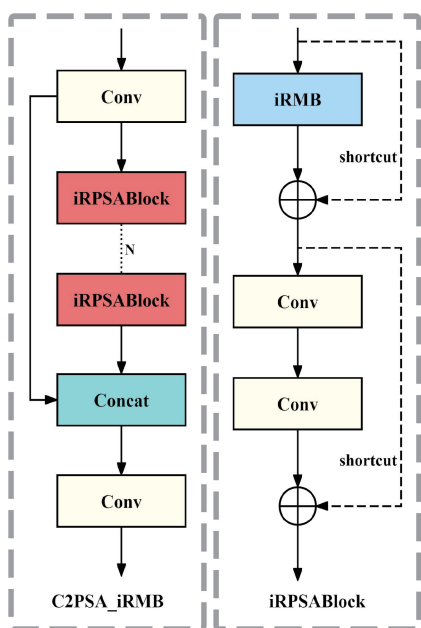


图6 C2PSA_iRMB 模块结构

Fig. 6 C2PSA_iRMB module structure diagram

程,但其较大的步长可能会导致边缘、纹理等细节特征信息的丢失,从而使提取的特征不够全面;并且传统卷积自身参数量较大,会造成模型计算负担加重。因此,模型使用传统卷积进行下采样的方式在水下生物检测中并不适用,因为水下生物中包含大量小目标,细节特征信息的缺失与高参数量的双重影响,会极大降低模型对水下小目标的检测性能。为解决上述问题,引入YOLOv9模型的ADown下采样模块替换YOLOv11模型中原有的传统卷积下采样模块,ADown下采样模块的具体结构如图7所示。

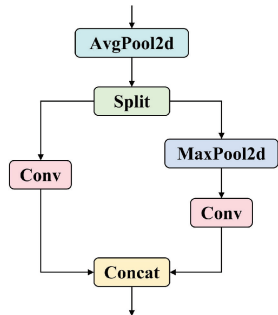


图7 ADown 下采样模块结构

Fig. 7 ADown downsampling module structure diagram

由图7可知,ADown下采样模块采用了双分支特征处理策略,具体实现流程如下:首先,通过二维平均池化对输入特征图进行下采样,将其空间尺寸减半;随后,特征图沿通道维度拆分为两部分,一部分直接应用 3×3 卷积操作,在提取空间特征的同时实现通道降维;另一部分则先通过最大平均池化提取主要特征信息,再利用 1×1

逐点卷积增加非线性变换表示并进一步压缩通道维度;最后,将这两部分处理后的特征图沿通道维度拼接,生成最终的输出张量。

与传统卷积下采样模块不同,ADown下采样模块凭借其独特的双分支结构设计,在下采样过程中能够极大地保留上下文语义信息并减少小目标特征信息的丢失。此外,ADown下采样模块采用了小尺寸卷积核,这使得它在保持强大特征表达能力的同时,有效控制了参数量和计算量。因此,本研究在YOLOv11模型中引入ADown下采样模块后,不仅提升了模型的训练和推理效率,还提高了水下生物检测的精度,同时进一步降低了模型的参数量和计算复杂度。

2.5 模块协同工作机理分析

YOLOv11-PiA模型的3大改进模块并非独立作用,而是通过“轻量化特征提取-多尺度特征增强-高效下采样整合”的递进式流程形成协同效应。

具体来说,C3k2_PConv模块的核心作用是轻量化特征提取,通过部分卷积PConv的选择性卷积策略,对输入特征图中 $1/4$ 的关键通道进行深层空间特征提取,其余通道直接保留浅层纹理信息。这种设计不仅降低了模型参数量与计算复杂度,还降低了输出特征图的通道冗余度,为后续C2PSA_iRMB模块的特征增强提供了高质量输入。在C2PSA_iRMB模块中,PSA注意力机制通过多尺度卷积核捕捉特征依赖关系,iRMB注意力机制通过扩展窗口多头自注意力实现全局-局部特征融合。输入特征包含大量冗余通道或无效背景信息会分散其注意力权重分配,降低特征增强效率。而C3k2_PConv模块输出的轻量化特征,能够使PSA注意力机制避免对无效通道的冗余计算,从而聚焦于水下生物的关键特征通道。与此同时,C3k2_PConv模块保留的浅层纹理特征也为iRMB注意力机制的局部特征建模提供了基础,并与自注意力捕捉的全局信息形成了互补,让特征增强过程更具针对性,最终提升模型对水下多尺度目标、形态相似目标的区分能力。

ADown模块的核心功能是高效下采样与特征整合,弥补了C3k2_PConv模块和C2PSA_iRMB模块在特征维度压缩与细节保留上的不足。在特征流转过程中,C3k2_PConv模块和C2PSA_iRMB模块侧重特征提取与增强,但未解决下采样过程中的细节丢失问题。而ADown模块的双分支结构极大地保留了C3k2_PConv模块和C2PSA_iRMB模块提取的纹理细节特征和语义特征,有效避免了传统卷积下采样导致的边缘信息丢失。此外,ADown模块的通道压缩策略与C3k2_PConv模块和C2PSA_iRMB模块也形成了阶梯式适配。其中,C3k2_PConv模块将输入通道精简至 $1/4$,C2PSA_iRMB模块通过注意力机制动态调整通道权重,ADown模块则通过双

分支卷积进一步将通道维度压缩至 1/2。这种阶梯式的通道设计,既避免了特征维度变换导致的信息损失,又通过通道降维降低了计算开销,最终使模型在保持轻量化的同时,提高了检测精度。

综上,C3k2_PConv 模块的轻量化设计解决了水下资源受限问题,C2PSA_iRMB 模块的特征增强能力解决了水下复杂环境导致特征表达不足的问题,ADown 模块的高效下采样解决了细节丢失与计算量平衡的问题。三者相互适配与约束,使模型在参数量降低的前提下,检测精度与速度同步提升,最终实现模型在水下生物检测任务中精度、速度与轻量化的高效平衡。

3 实验

3.1 实验环境和参数设置

所有实验的硬件配置如下:CPU 为 Intel(R) Xeon(R) Gold 6430 (16 核),GPU 为 NVIDIA GeForce RTX 4090(24 GB 显存),操作系统为 Ubuntu 18.04,编程环境为 Python 3.9,深度学习框架选用 PyTorch 2.0.1,CUDA

版本为 11.8。

主要训练参数设置如下:批处理大小(batch-size)为 16,训练轮数(epochs)为 200,训练图像的输入尺寸(img-size)为 640×640 pixels,模型训练采用 SGD 优化器,初始学习率为 0.01,动量初始值为 0.937,权重衰减系数为 0.000 5,workers 为 16。

3.2 实验数据集

使用公开的 RUOD 水下生物数据集^[22]和 DUO 水下生物数据集^[23]。

与现有水下生物数据集不同,RUOD 数据集涵盖了常见的自然水下场景和形态各异的水下生物,数据集样本本如图 8 所示。该数据集共包含 14 000 张图片,涉及海胆(echinus)、海参(holothurian)、海星(starfish)、扇贝(scallop)、海龟(turtle)、鱼(fish)、潜水员(diver)、水母(jellyfish)、珊瑚(corals)和墨鱼(cuttlefish)10 类目标。按照 7:1:2 的比例对原始数据集进行随机划分,构建训练集、验证集与测试集。划分完成后,数据集中 9 800 张图片用于模型训练,1 400 张图片用于模型验证,2 800 张图片用于模型测试。



图 8 RUOD 数据集样本
Fig. 8 RUOD dataset sample diagram

DUO 数据集收集并重新标注了多年的 URPC 数据集^[24],该数据集中的图片具有对比度低、图像模糊等特点,如图 9 所示。该数据集共有 7 782 张图片,包含海参、海胆、扇贝和海星 4 个类别,其中训练集 6 671 张,测试集 1 111 张。本文在保持原始测试集不变的基础上,对原始训练集按照 8.5:1.5 的比例随机划分为新的训练集和验证集,划分之后的数据集中 5 671 张用于训练,1 000 张用于验证,1 111 张用

于测试。

3.3 评价指标

采用召回率(recall,R)、平均精度(average precision, AP)、平均精度均值(mAP)、模型参数量(model parameters, Params)、浮点运算量(giga floating-point operations per second, GFLOPs)、帧率(frames per second, FPS)以及模型大小(model size)等多个评价指标评估模型性能。以下是这些评价指标的具体定义:

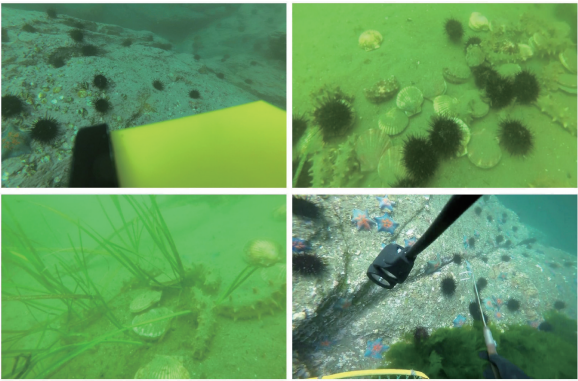


图 9 DUO 数据集样本图

Fig. 9 DUO dataset sample diagram

召回率定义为被模型正确检测出来的目标数量与所有目标数量的比值,计算公式为:

$$R = \frac{TP}{TP + FN}$$
(5)

式中: TP 表示被模型正确检测出来的目标数量; FN 表示没有被模型检测出来的目标数量,即漏检数量。

平均精度为 PR 曲线与坐标轴围成的面积,表示在多个不同召回率水平下的精度的平均值,计算公式为:

$$AP = \int_0^1 P(R) dR$$
(6)

平均精度均值(mAP)定义为所有类别平均精度的平均值,用于衡量模型在所有类别上的整体性能,计算公式为:

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}$$
(7)

式中: N 表示类别个数。其中 $mAP@0.5$ 是当 IoU 阈值取值为 0.5 得出的 mAP 值, $mAP@0.5:0.95$ 是当 IoU 阈值取值从 0.5 到 0.95 计算得出的 mAP 值。

模型参数量表示模型的复杂性和规模。参数较少意味着模型更轻量,内存需求更低,易于部署。

浮点运算量(GFLOPs)的单位是 10 亿次浮点运算,用于衡量模型推理时的计算量。浮点运算数越低意味着计算成本越低,其计算公式依赖于模型的具体架构和输入尺寸。

帧率是衡量模型实时处理能力的指标。在固定硬件条件下,FPS 值越高,表示模型的推理速度越快,越适用于对延迟敏感的应用场景。FPS 的计算通常需要在具体硬件上运行模型,并统计其平均每秒能处理的图像数量。

模型大小是指模型存储时所占用的空间,通常以 MB 为单位。模型大小直接影响模型的存储、传输和部署效率。

3.4 消融实验

为验证本文所提改进方法在水下生物检测任务中的有效性,本研究以 YOLOv11n 模型为基准,在 RUOD 数据集和 DUO 数据集上分别进行了消融实验,实验结果如表 1、2 所示。表中“√”代表使用对应的模块,“×”代表没有使用对应的模块。

表 1 基于 RUOD 数据集的消融实验
Table 1 Ablation experiments based on the RUOD dataset

实验	C3k2_ PCConv	C2PSA_ iRMB	ADown	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Params/(×10 ⁶)	浮点数/GFLOPs	帧率/fps	模型大小/ MB
1	×	×	×	75.5	82.8	56.6	2.58	6.3	88	5.5
2	√	×	×	76.7	84.2	58.8	2.28	6.1	91	4.8
3	×	√	×	77.2	83.9	58.6	2.60	6.3	89	5.5
4	×	×	√	76.0	83.7	58.1	2.10	5.1	91	4.5
5	√	√	×	77.8	85.1	60.2	2.29	6.1	93	4.9
6	√	×	√	78.2	85.2	60.3	1.79	4.9	95	3.9
7	×	√	√	77.8	84.6	59.4	2.12	5.2	95	4.6
8	√	√	√	78.6	85.6	60.9	1.81	4.9	97	3.9

表 1 的消融实验结果表明,在实验 2 中,当 C3k2_ PCConv 模块替换原模型的 C3k2 模块后,模型的 R 、 $mAP@0.5$ 、 $mAP@0.5:0.95$ 以及 FPS 分别提升了 1.2%、1.4%和 2.2%和 3 fps,且参数量和浮点运算数下降。这是因为 C3k2_ PCConv 模块通过部分卷积 PCConv 的选择性卷积策略,仅对 1/4 的关键特征通道执行了深层卷积操作,这在剔除冗余背景信息的同时又保留了水下生物的核心轮廓与纹理特征,既降低了计算开销,又避免了浅层

特征的过度损耗,实现了高精度与轻量化的平衡。由实验 3 结果可知,当结合了 iRMB 注意力机制的 C2PSA_ iRMB 模块替换原模型的 C2PSA 模块后,模型的 R 和 mAP 均有所提升。这一性能变化的核心原因在于,iRMB 注意力机制的扩展窗口自注意力可捕捉小目标的局部关键特征,而 PSA 注意力机制的多尺度卷积核可进一步强化跨尺度特征响应,这有效解决了原始 C2PSA 模块对小目标特征提取不充分的问题。同时,

iRMB 注意力机制通过深度可分离卷积和残差连接相结合的方式,不仅有效抑制了背景噪声干扰,还使模型更专注于水下生物目标的语义特征,从而减少了因复杂背景导致的误检现象。实验 4 结果表明,在 YOLOv11n 模型中引入 ADown 下采样模块后,模型的参数量和浮点运算数大幅下降且检测精度有所提升。这是因为 ADown 下采样模块采用了小卷积核进行逐点卷积,参数量远低于传统卷积下采样,同时其通道压缩策略也有效降低了模型的计算开销。此外,该模块的双分支下采样结构极大保留了边缘细节特征,这解决

了水下生物轮廓特征丢失的问题,降低了边界框回归误差。通过实验 5~7 结果可知,3 个改进模块的任意两者结合均能进一步提升模型的性能。

在实验 8 中,融合了 3 个改进模块的 YOLOv11-PiA 模型在所有性能评价指标上相较于 YOLOv11n 模型均有所提升。具体而言, R 提升了 3.1%、 $mAP@0.5$ 提升了 2.8%、 $mAP@0.5:0.95$ 提升了 4.3%、FPS 提升了 9 fps,且模型的参数量、浮点运算数和模型大小分别降低了 29.8%、22.2%和 29.1%。综上所述,本文提出的改进方法有效提升了模型在水下生物检测任务中的表现。

表 2 基于 DUO 数据集的消融实验

Table 2 Ablation experiments based on the DUO dataset

实验	C3k2_ PConv	C2PSA_ iRMB	ADown	$R/\%$	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	Params/ $(\times 10^6)$	浮点数/ GFLOPs	帧率/fps	模型大小/ MB
1	×	×	×	71.8	80.3	58.5	2.58	6.3	103	5.5
2	√	×	×	73.7	81.2	60.6	2.27	6.0	109	4.8
3	√	√	×	75.0	81.9	60.7	2.29	6.1	114	4.9
4	√	√	√	74.5	82.6	62.1	1.81	4.9	116	3.9

由表 2 的实验 2 结果可知,将 C3k2_PConv 模块替换 C3k2 模块后,模型的 R 、 $mAP@0.5$ 、 $mAP@0.5:0.95$ 和 FPS 分别提升了 1.9%、0.9%、2.1%和 6 fps,且模型的参数量和计算量降低。实验 3 在实验 2 的基础上引入 C2PSA_iRMB 模块替换 C2PSA 模块,模型的检测精度和速度得到了进一步提升。

实验 4 结果表明,当 C3k2_PConv 模块、C2PSA_iRMB 模块和 ADown 下采样模块三者协同作用时, YOLOv11-PiA 模型在各项评价指标上均优于 YOLOv11n 原模型,其中 R 提升了 2.7%, $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提升了 2.3%和 3.6%,FPS

提升了 13 fps,同时模型的参数量和计算复杂度降低,这再次证明了本研究改进方法的有效性和优越性。

3.5 模型对比实验

为进一步验证 YOLOv11-PiA 模型在水下生物目标检测任务中的优越性,本研究选取了当下流行的目标检测模型 YOLOv3-tiny^[25]、YOLOv5n^[26]、YOLOv6n^[27]、YOLOv8n^[28]、YOLOv9t^[16]、YOLOv10n^[29]、RTDETR-L^[30]、YOLOv12n^[31]以及部分现有水下生物检测模型在相同的实验环境和数据集下与本文所提出的 YOLOv11-PiA 模型进行了对比实验,实验结果如表 3、4 所示。

表 3 基于 RUOD 数据集的模型对比实验

Table 3 Model comparison experiments based on the RUOD dataset

模型	$R/\%$	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	Params/ $(\times 10^6)$	浮点数/GFLOPs	帧率/fps	模型大小/MB
YOLOv3-tiny	72.0	79.7	51.0	9.52	14.3	68	19.2
YOLOv5n	74.2	81.9	54.6	2.18	5.8	85	4.7
YOLOv6n	74.1	81.2	55.7	4.16	11.5	72	8.6
YOLOv8n	74.7	82.7	56.1	2.69	6.8	84	5.6
YOLOv9t	74.6	81.8	56.2	1.73	6.4	79	4.2
YOLOv10n	73.1	79.6	53.9	2.70	8.2	88	5.8
RTDETR-L	73.3	80.5	55.4	32.00	103.5	35	66.2
YOLOv11n	75.5	82.8	56.6	2.58	6.3	88	5.5
YOLOv12n	74.4	81.5	55.3	2.56	6.3	85	5.5
文献[9]	74.3	81.2	55.8	5.20	7.7	78	20.4
文献[11]	75.1	83.5	57.3	1.83	7.2	87	4.3
文献[12]	75.7	83.2	56.9	3.38	6.9	83	6.1
YOLOv11-PiA	78.6	85.6	60.9	1.81	4.9	97	3.9

表 3 的实验结果显示,虽然 YOLOv11-PiA 的参数量略高于 YOLOv9t,但仍符合轻量化标准。更重要的是, YOLOv9t 的召回率 R 、mAP 值和 FPS 远低于 YOLOv11-PiA,显然无法满足对检测精度和速度要求较高的水下生物检测应用场景。

此外,与 YOLOv3-tiny、YOLOv5n、YOLOv6n、YOLOv8n、YOLOv9t、YOLOv10n、RTDETR-L、YOLOv12n、文献[9,11-12]算法相比,YOLOv11-PiA 的 R 分别高出 6.6%、4.4%、4.5%、3.9%、4.0%、5.5%、5.3%、4.2%、4.3%、3.5%和 2.9%;mAP@0.5 分别高出 5.9%、3.7%、4.4%、2.9%、3.8%、6.0%、5.1%、4.1%、4.4%、2.1%和 2.4%;mAP@0.5:0.95 分别高出 9.9%、6.3%、5.2%、4.8%、4.7%、7.0%、5.5%、5.6%、5.1%、3.6%和 4.0%;

表 4 基于 DUO 数据集的模型对比实验

Table 4 Model comparison experiments based on the DUO dataset

模型	$R/\%$	mAP@0.5/%	mAP@0.5:0.95/%	Params/($\times 10^6$)	浮点数/GFLOPs	帧率/fps	模型大小/MB
YOLOv3-tiny	67.6	76.9	53.1	9.52	14.3	70	19.2
YOLOv5n	69.8	77.6	56.0	2.18	5.8	95	4.7
YOLOv6n	70.8	78.3	57.9	4.16	11.5	78	8.6
YOLOv8n	71.3	80.1	59.0	2.69	6.8	104	5.6
YOLOv9t	70.6	79.8	59.6	1.73	6.4	87	4.2
YOLOv10n	71.3	78.9	57.7	2.70	8.2	100	5.8
RTDETR-L	62.7	66.8	47.0	31.99	103.4	38	66.2
YOLOv11n	71.8	80.3	58.5	2.58	6.3	103	5.5
YOLOv12n	69.9	78.2	56.4	2.56	6.3	105	5.5
文献[9]	70.5	78.9	56.2	5.20	7.7	82	20.4
文献[11]	72.4	80.5	58.1	1.83	7.2	107	4.3
文献[12]	73.3	81.2	60.4	3.38	6.9	102	6.1
YOLOv11-PiA	74.5	82.6	62.1	1.81	4.9	116	3.9

3.6 不同注意力机制改进 C2PSA 模块对比实验

为了验证本研究在 C2PSA 模块中嵌入 iRMB 注意力机制的有效性与优越性,选取了 SEAM^[32]、iEMA^[33]、TripletAttention^[34]和 BiFormer^[35]4 种注意力机制,在与 iRMB 注意力机制相同的嵌入位置对 C2PSA 模块进行了改进,并在 RUOD 数据集上开展了对比实验,实验结果如

FPS 分别高出 29、12、25、13、18、9、62、12、9、10 和 14 fps,且 YOLOv11-PiA 模型的浮点运算量最低、和模型最小,这意味着 YOLOv11-PiA 模型更易于部署。综上所述,本文提出的 YOLOv11-PiA 模型在水下生物检测任务中实现了精度与轻量化的高效平衡,其综合性能显著优于现有主流目标检测模型。

根据表 4 基于 DUO 数据集的模型对比实验结果可知,虽然 YOLOv11-PiA 模型的参数量并非最低,但仍符合轻量化设计要求。此外,相较于其他目标检测模型, YOLOv11-PiA 模型在 DUO 数据集上的检测精度和速度均表现最佳,这也再次证明本研究改进方法的有效性和 YOLOv11-PiA 模型在水下生物目标检测任务中的优越性。

表 5 不同注意力机制改进 C2PSA 模块对比实验

Table 5 Comparative experiments on improving the C2PSA module with different attention mechanisms

注意力机制	$R/\%$	mAP@0.5/%	mAP@0.5:0.95/%	帧率/fps
SEAM	76.0	83.1	57.5	86
iEMA	76.6	83.4	57.4	85
TripletAttention	75.9	83.5	57.5	87
BiFormer	76.3	83.6	57.9	89
iRMB	77.2	83.9	58.6	89

表 5 所示。

根据表 5 实验结果可知,在 C2PSA 模块中嵌入 iRMB 注意力机制后,模型在 RUOD 数据集上的性能表现显著优于 SEAM、iEMA、TripletAttention 和 BiFormer4 种对比机制,充分验证了 C2PSA_iRMB 模块的有效性和优越性。

3.7 可视化分析

在目标检测算法中,精确率-召回率(P-R)曲线是评估算法性能的重要指标。图 10 为 YOLOv11n 模型与

YOLOv11-PiA 模型在 RUOD 数据集(图 10(a))和 DUO 数据集(图 10(b))上的 P-R 曲线对比结果。如图 10 所示,YOLOv11-PiA 模型的 PR 曲线始终位于 YOLOv11n

模型的上方,且其曲线所围成的面积更大。这一结果表明,YOLOv11-PiA 模型在水下生物检测任务中具有更高

的精确率和召回率,充分验证了本文改进方法的有效性和优越性。

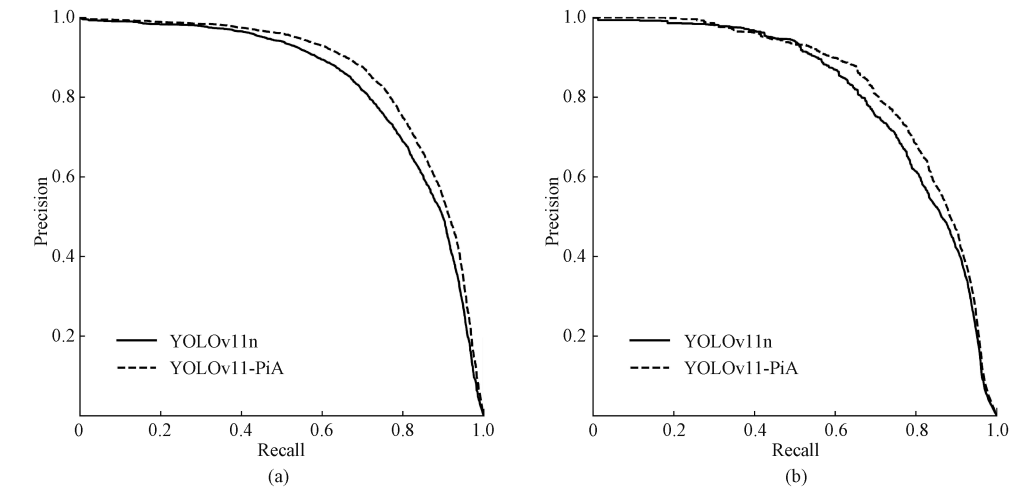


图 10 P-R 曲线对比图
Fig. 10 P-R curve comparison diagram

此外,本研究对 YOLOv11n 模型与 YOLOv11-PiA 模型在训练过程中的 $mAP@0.5$ 变化进行了可视化对比,

如图 11 所示。图 11(a) 为 RUOD 数据集的 $mAP@0.5$ 对比,图 11(b) 为 DUO 数据集的 $mAP@0.5$ 对比。

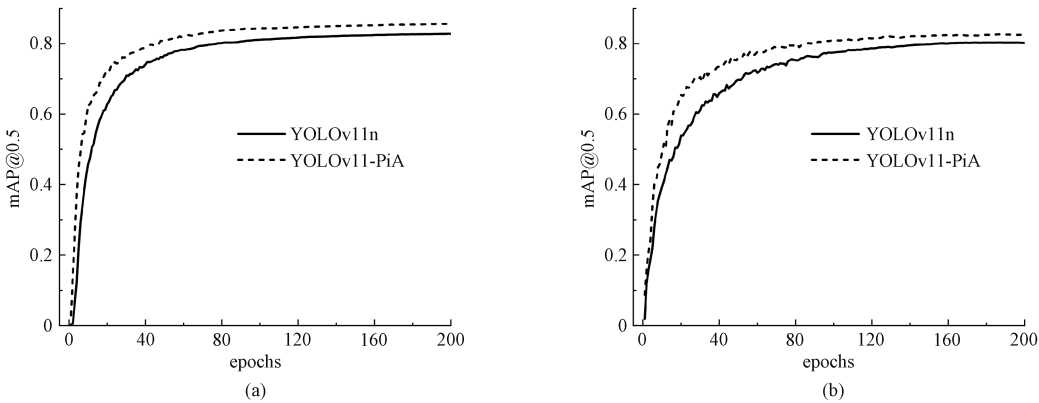


图 11 $mAP@0.5$ 对比
Fig. 11 $mAP@0.5$ comparison diagram

由图 11 可知,随着训练轮次的增加,两个模型的 $mAP@0.5$ 均呈现上升趋势,并在 160 轮后逐步趋于稳定。而在整个训练过程中,YOLOv11-PiA 模型的 $mAP@0.5$ 值始终高于 YOLOv11n 模型,这一结果表明,本文提出的改进方法能够有效提升模型在水下生物检测任务中的精度。

最后,为了直观展示改进的有效性和优越性,从 RUOD 测试集中选取了 4 张图片,用于对比 YOLOv11n 模型与 YOLOv11-PiA 模型的检测效果。如图 12 所示,图 12(a1)~(d1) 为 YOLOv11n 模型的检测结果,图 12(a2)~(d2) 为 YOLOv11-PiA 模型的检测结果。

YOLOv11-PiA 在检测效果上显著提升。图 12(a) 中,YOLOv11n 模型将墨鱼的触须误检为墨鱼,且在图片右上方也出现了将其他生物误检为鱼的现象。这些误检现象表明,原模型在相似水下生物的细粒度分类能力上存在不足,容易对具有相似形态特征的水下生物产生混淆。相比之下,改进后的 YOLOv11-PiA 模型不仅准确识别出所有目标生物,完全避免了误检情况,同时还显著提升了检测结果的置信度。图 12(b) 中,YOLOv11n 模型未能检测出图片中与礁石颜色相似的海龟,出现了漏检现象。这表明原模型在复杂水下场景中对特定目标的检测能力存在不足,特别是当目标被遮挡或与背景相似时更容易出现漏检。而本文提出的 YOLOv11-PiA 模型则成功检

通过对比 4 组图片可以看出,改进后的模型

测出了所有目标,有效减少了漏检情况的发生。图 12(c)为模型在水下模糊场景的检测结果。在这种场景下,原模型在图片左下方出现了误检,并且整体检测置信度偏低。而 YOLOv11-PiA 模型不仅没有出现误检现象,还在保持高召回率的同时提高了检测置信度,这表明本文的改进方法有效提升了模型在水下模糊场景的识别能力。图 12(d)中,目标数量众多且以小目标为主,这是水下生物数据集的常见特征。面对这种复杂情况,YOLOv11n 模型出现了严重的误检现象,且对小目标的检测置信度较低,这在实际应用中可能会带来严重后果。然而,经过改进后的 YOLOv11-PiA 模型能够基本准确地检测到每一个目标,显著改善了误检现象,并且提高了对小目标的检测置信度,这说明 YOLOv11-PiA 模型相较于原模型更适用于水下多目标场景。

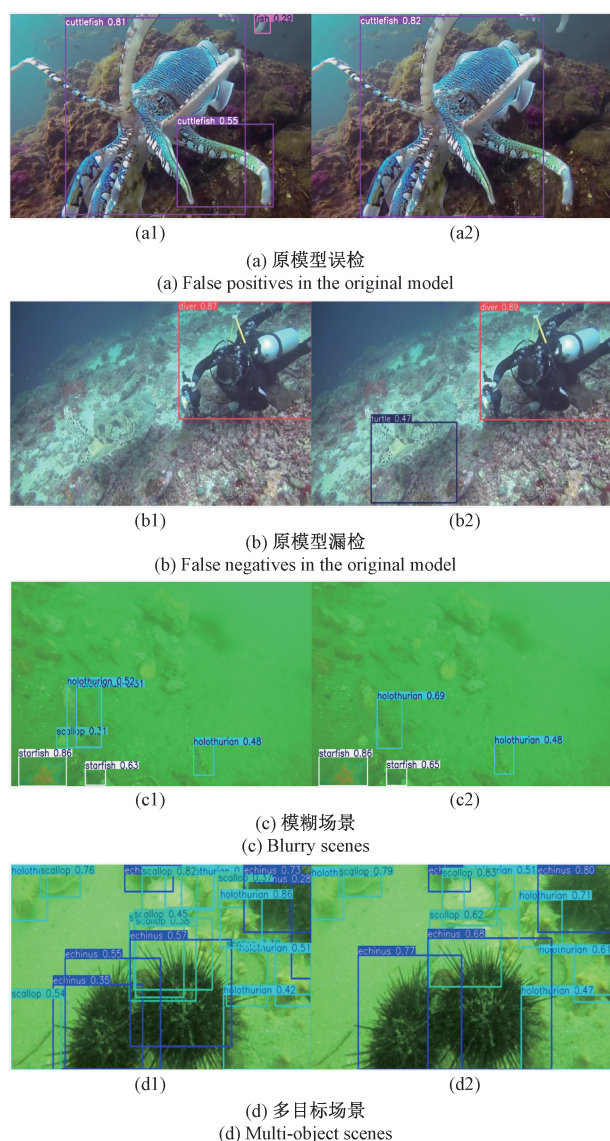


图 12 模型检测结果对比

Fig. 12 Model detection result comparison diagram

综合图 12 的 4 组对比效果得知,改进后的 YOLOv11-PiA 模型相较于 YOLOv11n 原模型在检测效果上有着明显的提升,再次验证了本文所提改进方法的有效性和 YOLOv11-PiA 模型在水下生物检测任务中的优越性。

4 结 论

为应对水下环境的高度复杂性以及解决现有水下生物目标检测模型参数量较大的问题,本文提出了一种改进 YOLOv11 的轻量化水下生物目标检测算法 YOLOv11-PiA。该算法通过采用部分卷积 PConv 对 C3k2 模块进行改进,构建了 C3k2_PConv 模块,在减少模型的参数量和计算量的同时提高了检测精度和速度。针对水下复杂环境中的噪声问题,引入倒置残差块注意力机制 iRMB 对 C2PSA 模块进行改进,提出了 C2PSA_iRMB 模块,有效增强了模型的表征能力,进而提升了检测精度。此外,将基线模型中的传统卷积下采样模块替换为 ADown 下采样模块,进一步降低了模型的参数量并再次提高了检测效率。实验证明,在水下生物数据集 RUOD 和 DUO 上,YOLOv11-PiA 在检测精度和速度、模型参数量、计算复杂度等多个指标上均优于原模型和其他先进的目标检测模型,满足了轻量化和实时性需求,为水下生物目标检测提供了一种高效的解决方案。然而,本研究仍存在一定不足,模型在面对形态特殊的水下生物目标时,检测框定位精度易出现偏差,尤其是在生物躯体弯曲或部分躯体被遮挡时,边界框回归误差会显著增大;并且在极端水下环境(如高浑浊度水体、强水流扰动导致画面模糊)中,模型对目标特征的提取能力会明显下降,易出现漏检。这些不足在实际应用场景中可能会对模型的检测效果产生一定影响。因此,未来的研究将针对这些不足进行改进,例如设计基于形态学特征增强的边界框修正机制以提升对弯曲、遮挡状态下目标的定位精度;构建多模态数据融合模块,利用声呐数据在浑浊环境中的穿透性优势弥补光学图像的特征损失,进而增强模型在极端水下环境中的鲁棒性。

参考文献

- [1] 彭伟,厉运周,高艳波,等.我国海洋观测仪器设备体系化发展研究[J].仪器仪表学报,2023,44(12): 88-100.
PENG W, LI Y ZH, GAO Y B, et al. Research on systematic development of marine observation instruments and equipment in China [J]. Chinese Journal of Scientific Instrument, 2023, 44(12): 88-100.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and

- semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multiBox detector [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016: 21-37.
- [4] 王宁, 智敏. 深度学习下的单阶段通用目标检测算法研究综述[J]. 计算机科学与探索, 2025, 19(5): 1115-1140.
- WANG N, ZHI M. A survey of single-stage general object detection algorithms based on deep learning[J]. Journal of Frontiers of Computer Science and Technology, 2025, 19(5): 1115-1140.
- [5] SHI P, XU X, NI J, et al. Underwater biological detection algorithm based on improved Faster-RCNN[J]. Water, 2021, 13(17): 2420.
- [6] HUANG A, ZHONG G, LI H, et al. Underwater object detection using restructured SSD[C]. CAAI International Conference on Artificial Intelligence. Cham: Springer Nature Switzerland, 2022: 526-537.
- [7] 史朋飞, 韩松, 倪建军, 等. 结合数据增强和改进 YOLOv4 的水下目标检测算法[J]. 电子测量与仪器学报, 2022, 36(3): 113-121.
- SHI P F, HAN S, NI J J, et al. Underwater object detection algorithm combining data augmentation and improved YOLOv4 [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 113-121.
- [8] 张银胜, 杨宇龙, 胡宇翔, 等. 改进视觉注意力网络的水下目标检测算法[J]. 国外电子测量技术, 2023, 42(12): 132-143.
- ZHANG Y SH, YANG Y L, HU Y X, et al. Underwater object detection algorithm with improved visual attention network [J]. Foreign Electronic Measurement Technology, 2023, 42(12): 132-143.
- [9] 周新, 张春堂, 樊春玲. 基于 YOLOv5_PGS 的轻量级水下生物识别目标检测[J]. 电子测量技术, 2023, 46(21): 168-175.
- ZHOU X, ZHANG CH T, FAN CH L. Lightweight underwater biological recognition object detection based on YOLOv5 _ PGS [J]. Electronic Measurement Technology, 2023, 46(21): 168-175.
- [10] YI W, WANG B. Research on underwater small target detection algorithm based on improved YOLOv7 [J]. IEEE Access, 2023, 11: 66818-66827.
- [11] 苗力恒, 田莹. 改进 YOLOv8n 的多尺度轻量化水下目标检测[J]. 电子测量与仪器学报, 2025, 39(4): 141-151.
- MIAO L H, TIAN Y. Multi-scale lightweight underwater object detection with improved YOLOv8n[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(4): 141-151.
- [12] PAN W, CHEN J, LV B, et al. Lightweight marine biodetection model based on improved YOLOv10 [J]. Alexandria Engineering Journal, 2025, 119: 379-390.
- [13] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements [J]. ArXiv preprint arXiv:2410.17725, 2024.
- [14] CHEN J, KAO S, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [15] ZHANG J, LI X, LI J, et al. Rethinking mobile block for efficient attention-based models[C]. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2023: 1389-1400.
- [16] WANG C Y, YE H I, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 1-21.
- [17] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context [C]. Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, 2014: 740-755.
- [18] ZHANG H, ZU K, LU J, et al. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network [C]. Proceedings of the Asian Conference on Computer Vision, 2022: 1161-1177.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [20] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 1-15.
- [22] FU C, LIU R, FAN X, et al. Rethinking general underwater object detection: Datasets, challenges, and solutions[J]. Neurocomputing, 2023, 517: 243-256.
- [23] LIU C, LI H, WANG S, et al. A dataset and benchmark of underwater object detection for robot picking[C]. 2021

- IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2021: 1-6.
- [24] 罗逸豪,刘奇佩,张吟,等. 基于深度学习的水下图像目标检测综述[J]. 电子与信息学报,2023,45(10): 3468-3482.
- LUO Y H, LIU Q P, ZHANG Y, et al. A survey of underwater image object detection based on deep learning[J]. Journal of Electronics & Information Technology, 2023, 45(10): 3468-3482.
- [25] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [26] KHANAM R, HUSSAIN M. What is YOLOv5: a deep look into the internal features of the popular object detector[J]. ArXiv preprint arXiv:2407.20892, 2024.
- [27] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.
- [28] SOHAN M, SAI RAM T, RAMI REDDY C V. A review on YOLOv8 and its advancements [C]. International Conference on Data Intelligence and Cognitive Informatics. Springer, Singapore, 2024: 529-545.
- [29] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [30] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [31] TIAN Y, YE Q, DOERMANN D. YOLOv12: attention-centric real-time object detectors [J]. Advances in Neural Information Processing Systems, 2026, 38: 78433-78457.
- [32] YU Z, HUANG H, CHEN W, et al. YOLO-FaceV2: A scale and occlusion aware face detector [J]. Pattern Recognition, 2024, 155: 110714.
- [33] 朱强军,胡斌. 基于 IEMA 和混合小目标的 YOLOv8s 交通标志检测方法[J]. 湖北民族大学学报(自然科学版),2025,43(2):202-209.
- ZHU Q J, HU B. Traffic sign detection method based on IEMA and hybrid small objects using YOLOv8s [J]. Journal of Hubei Minzu University (Natural Science Edition), 2025, 43(2): 202-209.
- [34] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3139-3148.
- [35] ZHU L, WANG X, KE Z, et al. BiFormer: Vision transformer with bi-level routing attention [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.

作者简介



郭文豪,2023 年于郑州科技学院获得学士学位,现为内蒙古科技大学硕士研究生,主要研究方向为水下生物目标检测。

E-mail:2248707737@qq.com

Guo Wenhao received his B. Sc. degree from Zhengzhou University of Science and Technology in 2023. Now he is a M. Sc. candidate at Inner Mongolia University of Science and Technology. His main research interests include underwater biological target detection.



郝斌(通信作者),2007 年于浙江大学获得学士学位,2010 年于浙江大学获得硕士学位,2023 年于清华大学获得博士学位,现为内蒙古科技大学讲师,主要研究方向为目标检测、推荐系统。

E-mail:haobin9527@qq.com

Hao Bin (Corresponding author) received his B. Sc. degree from Zhejiang University in 2007, M. Sc. degree from Zhejiang University in 2010, and Ph. D. degree from Tsinghua University in 2023, respectively. Now he is a lecturer at Inner Mongolia University of Science and Technology. His main research interests include target detection and recommendation systems.