文章编号:1674-2869(2026)01-0082-08

改进YOLOv11n的复杂水下场景目标检测模型

潘文丽,徐国庆*,王小甜,乐文波
武汉工程大学计算机科学与工程学院、人工智能学院,湖北 武汉 430205

摘 要:为提升海洋探测开发中水下目标检测的准确率,针对复杂水下环境中目标轮廓模糊、生物聚集遮挡及多尺度目标检测精度不足等问题,提出了一种改进YOLOv11n的复杂水下场景目标检测模型。首先,在基线模型YOLOv11n的骨干网络中引入轻量化混合架构EfficientFormerV2,通过Transformer的多阶段协同设计,有效捕获局部细节与全局语义信息,缓解低分辨率及特征损失对检测性能的影响;其次,设计自适应上下文融合模块,采用局部和全局双分支结构并结合可学习权重机制,实现跨层特征的高效对齐与融合,增强模型在复杂水下场景下的目标分辨能力;最后,引入动态检测头,通过尺度感知、空间聚焦和任务解耦的联合优化,提升对多尺度水下生物目标的检测精度。实验结果表明,在RUOD数据集和DUO数据集上,改进后的模型相较于YOLOv11n,交并比阈值为50%时的平均精度均值分别提升了1.3%和1.8%,模型参数量仅增加3.4×10^6,整体性能超越现有主流方法。该模型有效克服了水下特征退化及多尺度检测精度不足的问题,提升了复杂水下环境下的目标识别精度,为海洋资源勘探、沉船定位和生态监测提供高效解决方案。

关键词:水下目标检测;YOLOv11n;特征融合;注意力机制
中图分类号:TP391.4 **文献标识码**:A **DOI**:10.19843/j.cnki.CN42-1779/TQ.202410020

An enhanced aquatic detection-YOLOv11n model for target detection in complex underwater scenes

PAN Wenli, *XU Guoqing**, *WANG Xiaotian*, *YUE Wenbo*
School of Computer Science and Engineering, School of Artificial Intelligence, Wuhan Institute of Technology, Wuhan 430205, China

Abstract: To address challenges such as blurred target contours, occlusion by biological aggregations, and insufficient multi-scale target detection accuracy in complex underwater environments, in this paper, we proposed an enhanced aquatic detection-YOLOv11n (EAD-YOLOv11n) model for target detection in complex underwater scenes. First, a lightweight hybrid architecture, EfficientFormerV2, was integrated into the backbone network of the baseline YOLOv11n model. Leveraging a multi-stage collaborative design of Transformer, it effectively captured both local details and global semantic information, mitigating the impact of low resolution and feature loss on detection performance. Second, an adaptive context fusion module (ACFM) was designed. Employing a dual-branch structure for local and global features combined with a learnable weight mechanism, it achieved efficient cross-layer feature alignment and fusion, enhancing the model's capability to distinguish targets in complex underwater scenes. Finally, a dynamic head (DyHead) was introduced. Through joint optimization incorporating scale awareness, spatial focusing, and task

收稿日期:2024-10-28
基金项目:武汉大学测绘遥感信息工程国家重点实验室开放基金(21E01)
作者简介:潘文丽,硕士研究生。Email:3335676790@qq.com
***通信作者**:徐国庆,博士,副教授。Email:124148659@qq.com
引文格式:潘文丽,徐国庆,王小甜,等. 改进YOLOv11n的复杂水下场景目标检测模型[J]. 武汉工程大学学报,2026,48(1):82-89.

decoupling, it improved the model's accuracy for detecting multi-scale underwater biological targets. Experimental results demonstrated that, compared to the baseline YOLOv11n, the enhanced model achieved improvements of 1.3% and 1.8% in mean average precision at a 50% intersection over union threshold (mAP@50) on the RUOD and DUO datasets, respectively, with only a 3.4×10^6 parameter increase, and outperformed existing mainstream methods. The model proposed effectively overcomes underwater feature degradation and insufficient multi-scale detection accuracy, enhancing target recognition performance in complex underwater environments, thereby providing an efficient solution for marine resource exploration, shipwreck localization, and ecological monitoring.

Keywords: underwater target detection; YOLOv11n; feature fusion; attention mechanism

水下目标检测技术在海洋生物探测[1]、环境监测[2]和水下考古[3]等领域具有重要的应用价值，已成为计算机视觉领域的研究热点。然而，水下环境的特殊性给目标检测带来了巨大挑战。水体的吸收和散射效应导致水下图像普遍存在颜色失真、对比度降低和雾化等问题[4]。同时，复杂的背景干扰、显著的目标尺度差异以及频繁的生物遮挡现象，进一步增加了检测难度。传统的目标检测方法难以应对这些复杂情况[5]。近年来，随着深度学习技术的快速发展，基于深度学习的水下目标检测方法取得了显著进展，但在实际应用中仍面临诸多技术瓶颈。

当前主流方法主要基于单阶段检测框架，通过优化网络结构与特征融合机制提升检测性能。双阶段检测器，例如加速的区域卷积神经网络[6](faster region-based convolutional neural network, Faster R-CNN)、分割掩模的区域卷积神经网络[7](mask region-based convolutional neural network, Mask R-CNN)，虽具有较高检测精度，但其复杂的区域建议机制导致计算成本过高，难以满足实时性需求。相较之下，YOLO[8]系列、单点多盒探测[9](single shot multibox detector, SSD)等单阶段算法凭借高效的端到端检测框架，在速度与精度间取得更好平衡，逐渐成为水下检测的主要技术路线。

针对水下特殊环境，研究者提出了多种改进策略。陈宇梁等[10]提出了基于改进YOLOv5s的弱光水下生物目标检测模型YOLOv5s-underwater，通过自适应直方图均衡改善颜色失真，引入快速空间金字塔池化模块提升泛化能力，适配弱光水下环境。辛世澳等[11]采用轻量级网络改进了YOLOv7的主干，显著减少了模型的参数量和计算量，并通过使用双向特征金字塔网络进行多尺度特征融合，利用双向路由注意力视觉变换器提升了关键特征信息，改善水下目标特征模糊导致的检测问题。李培坤等[12]采用分组卷积和逐点卷积的思想改进YOLOv8n网络中的C2f模块，增加小目标检测层以提高对水下小目标的检测能力，并在颈部网络中引入轻量级上采样算子和注意力机制，提升模型对水下模糊目标的检测精度。这些方法虽在不同程度上提升了检测性能，但现有水下目标检测方法仍受限于轮廓模糊、多尺度特征融合不佳及固定检测头难以适应目标尺度变化等问题，检测精度不足。

为此，本研究基于YOLOv11n提出改进模型：采用EfficientFormerV2主干网络，通过注意力机制优化特征提取，增强小目标检测能力；设计自适应上下文融合模块(adaptive context fusion module, ACFM)，通过双分支动态融合多尺度特征，提升复杂环境鲁棒性；引入动态检测头(dynamic head, DyHead)，优化多尺度目标检测精度。改进模型旨在提高水下目标检测性能，为海洋勘探和生态监测提供高效解决方案。

1　YOLOv11n模型改进

YOLOv11n采用改进的CSPDarknet53作为骨干网络，使用C3K2模块替代传统的C2f模块，同时引入C2PSA模块增强多尺度特征提取能力。在颈部网络中，YOLOv11n结合了路径聚合网络与特征金字塔网络，通过融合深浅层特征提升特征表达的丰富性。此外，检测头采用解耦结构，并引入深度可分离卷积，以进一步减少计算开销。

水下环境中光照不均，生物目标尺度多样、形态复杂且背景杂乱，给目标检测带来了巨大挑战。为此，本研究提出基于YOLOv11n的增强型水下检测(enhanced aquatic detection-YOLOv11n, EAD-YOLOv11n)模型。采用EfficientFormerV2[13]替代原始特征提取网络，增强局部细节和全局语义理解，从而提高模型在低对比度、光照不均环境下的目标感知能力。主干网络的特征层P3、P4和P5提

供不同粒度信息，P3层用于捕捉小目标特征，P4层提供中等目标特征，P5层主要用于获取大尺度特征。通过ACFM模块将这3层特征与颈部网络的输出进行动态融合，以优化跨通道信息整合。此外，引入DyHead，通过动态空间定位、任务判别和多尺度特征融合，提高检测一致性和多尺度适应性，从而增强水下目标的检测精度和鲁棒性。EAD-YOLOv11n模型结构如图1所示。

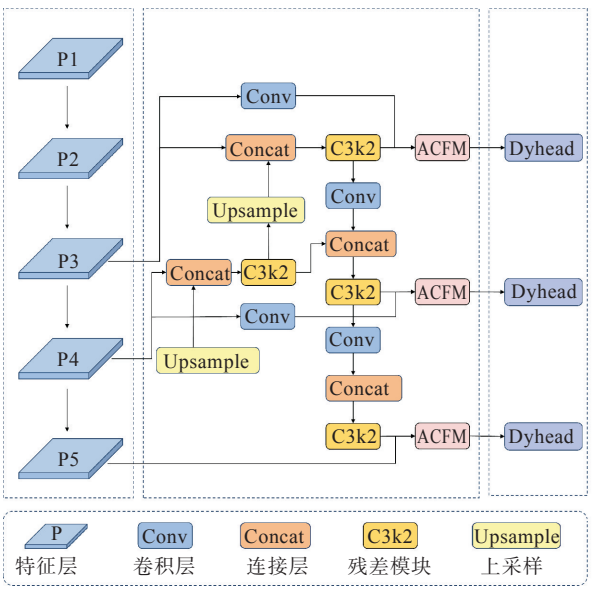

图1 网络总体结构

Fig. 1 Overall network structure

1.1 特征提取网络的改进

水下目标检测面临目标轮廓模糊、水体浑浊及设备算力受限等挑战，要求模型对边缘特征敏感且轻量化。传统注意力机制虽能通过特征聚焦抑制噪声干扰，但其参数冗余与复杂计算结构难以适配低算力场景。为此，本研究采用EfficientFormerV2作为新型主干网络，替代计算密集型的CSPDarknet53，通过纯Transformer架构与延迟优化策略实现精度和效率的平衡。

EfficientFormerV2采用四阶段分层式设计，通过空间分辨率递减与通道维度递增的协同优化，实现局部细节保持与全局上下文建模的动态平衡。EfficientFormerV2结构如图2所示，其中C_j为第j阶段的通道数，H、W分别为输入特征图的高和宽。

输入图像首先经过由2个3×3卷积构成的小内核卷积进行初步特征提取与下采样，替代了传统视觉转换器的非重叠补丁嵌入，有助于提升计算效率。前两个阶段的局部特征提取专注于高分辨率下的局部信息，采用相同的前馈神经网络

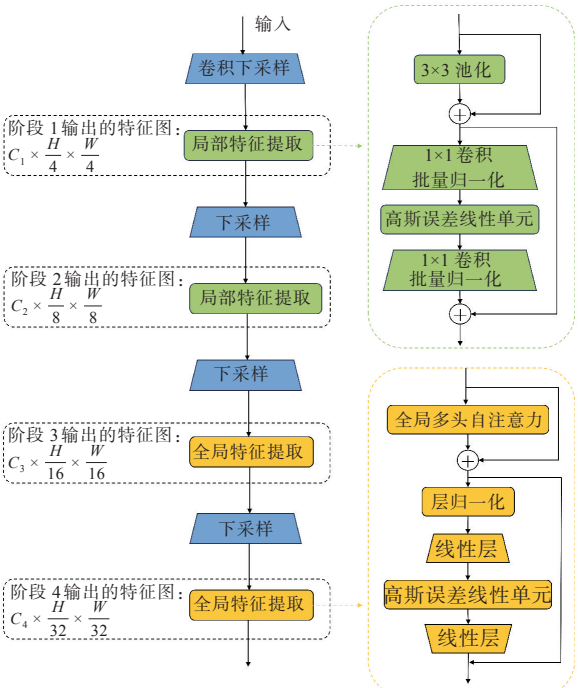

图2 EfficientFormerV2主干网络

Fig. 2 EfficientFormerV2 backbone network

（feedforward neural network，FNN）处理每层特征图，每个FNN之间均用残差连接，表达式如下：

$$\boldsymbol{X}_{i+1,j}^{B,C_j,\frac{H}{2^{j+1}},\frac{W}{2^{j+1}}}=\boldsymbol{S}_{i,j}\cdot F_{i,j}\left(\boldsymbol{X}_{i,j}\right)+\boldsymbol{X}_{i,j} \tag{1}$$

式中：$\boldsymbol{X}_{i,j}$是第i层第j阶段的特征，$j\in\{1,2,3,4\}$；B表示批次大小；$\boldsymbol{S}_{i,j}$表示第i层第j阶段的可学习尺度因子；$F_{i,j}$为第i层第j阶段的FNN的输出特征，FNN的宽度由通道数C_j决定。

后两个阶段的全局特征提取采用多头自注意力（multi-head self-attention，MHSA）与FNN的混合设计，兼顾了全局上下文建模与局部细节的保留，表达式如下：

$$\boldsymbol{X}_{i+1,j}^{B,C_j,\frac{H}{2^{j+1}},\frac{W}{2^{j+1}}}=\boldsymbol{S}_{i,j}\cdot M\left(\boldsymbol{P}\left(\boldsymbol{X}_{i,j}\right)\right)+\boldsymbol{X}_{i,j} \tag{2}$$

$$M\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=\sigma\left(\boldsymbol{Q}\cdot\boldsymbol{K}^{\mathrm{T}}+\boldsymbol{ab}\right)\cdot\boldsymbol{V} \tag{3}$$

式中：M表示MHSA模块；σ为softmax归一化函数；$\boldsymbol{Q}$、$\boldsymbol{K}$和$\boldsymbol{V}$分别表示通过线性层$\boldsymbol{P}\left(\boldsymbol{X}_{i,j}\right)$从输入特征中投影得到的查询、键和值；$\boldsymbol{a}$和$\boldsymbol{b}$分别是与$\boldsymbol{Q}$和$\boldsymbol{K}$有关的可学习位置编码偏差，用于增强注意力机制的位置感知能力，帮助模型更好地理解序列中的位置关系。

1.2 特征融合网络的改进

水下目标检测受光衰减和散射影响，导致目标边界模糊且易与复杂背景混淆。YOLOv11n通过多尺度特征融合整合浅层位置信息与深层语义

特征，但尺度间特征平衡不足，感受野失配常导致小目标细节丢失或误检漏检。内容引导注意力融合[14]（content-guided attention fusion，CGAFusion）模块通过生成空间不变图并结合通道与空间注意力权重，增强特征通道间交互，有效缓解这一问题。然而，在复杂背景中，其动态特征调整能力不足。为此，本研究基于卷积与注意力融合模块[15]（convolution and attention fusion module，CAFM），提出 ACFM 模块。ACFM 通过局部与全局双分支协同融合策略，增强浅层与深层特征的互补性，提升跨尺度信息交互效率，有效应对水下图像的复杂场景。ACFM 模块结构如图 3 所示，其中 w 表示权重。

首先，子模块 CAFM 接收来自主干网络的浅层特征和深层特征，利用特征调制的空间权重，动态调整各特征的重要性，从而增强关键目标的表征能力并抑制复杂背景的干扰。其次，ACFM 通过跳跃连接将输入特征与输出特征融合，有效缓解梯度消失问题，优化模型训练过程，并提高特征传递效率。最后，融合后的特征通过 1×1 卷积层映射，生成最终的输出特征。

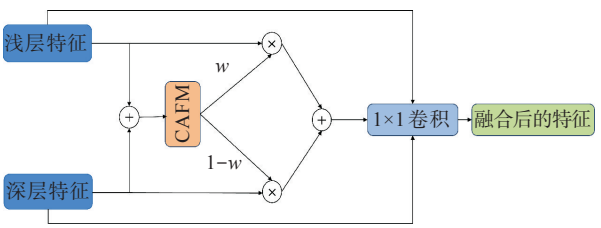

图 3　ACFM 模块

Fig. 3　ACFM module

ACFM 模块基于 CAFM 双分支结构实现，通过局部与全局特征协同融合增强了模型对不同空间位置和尺度特征的感知能力，并改善了跨通道信息的整合。CAFM 的具体结构如图 4 所示。

在局部分支中，为增强跨通道交互，首先通过 1×1 卷积调整通道数。随后执行通道混洗操作，将输入张量分组以提高信息融合度。各组内应用深度可分离卷积提取局部特征，组间输出沿通道维度串联。最后通过 3×3×3 卷积进一步提取特征。表达式如下：

$$\boldsymbol{F}_{\mathrm{Conv}} = C_{3\times3\times3}\left(R\left(C_{1\times1}(\boldsymbol{Y})\right)\right) \tag{4}$$

式中：$\boldsymbol{F}_{\mathrm{Conv}}$ 表示局部分支的输出特征图，C 表示卷积操作，R 表示通道混洗操作，$\boldsymbol{Y}$ 为输入特征图。

在全局分支中，通过 1×1 卷积和 3×3 深度卷积分别生成 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，得到 3 个形状为 $H \times W \times C$ 的张量，对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 进行重塑，然后通过交互计算得到全局注意力图，表达式如下：

$$\boldsymbol{F}_{\mathrm{att}} = C_{1\times1}A(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) + \boldsymbol{Y} \tag{5}$$

$$A(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \boldsymbol{V}\cdot\sigma\left(\frac{\boldsymbol{QK}}{\boldsymbol{\alpha}}\right) \tag{6}$$

式中：$\boldsymbol{F}_{\mathrm{att}}$ 表示全局分支的输出特征图；A 为注意力机制；$\boldsymbol{\alpha}$ 为可学习的缩放参数，在激活函数 σ 之前控制 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的矩阵乘法的幅度。

最后，CAFM 的输出 $\boldsymbol{F}_{\mathrm{out}}$ 表达式为：

$$\boldsymbol{F}_{\mathrm{out}} = \boldsymbol{F}_{\mathrm{att}} + \boldsymbol{F}_{\mathrm{Conv}} \tag{7}$$

通过局部与全局的双分支协同特征融合，ACFM 模块能够有效提升特征分布均匀性，改善跨通道信息整合问题，从而增强 YOLOv11n 的检测能力和性能。

1.3　检测头的改进

YOLOv11n 原有的检测头采用单尺度预测结构，未能充分利用多尺度特征，导致在复杂水下场景中对多尺度目标的检测精度不足。其次，传统

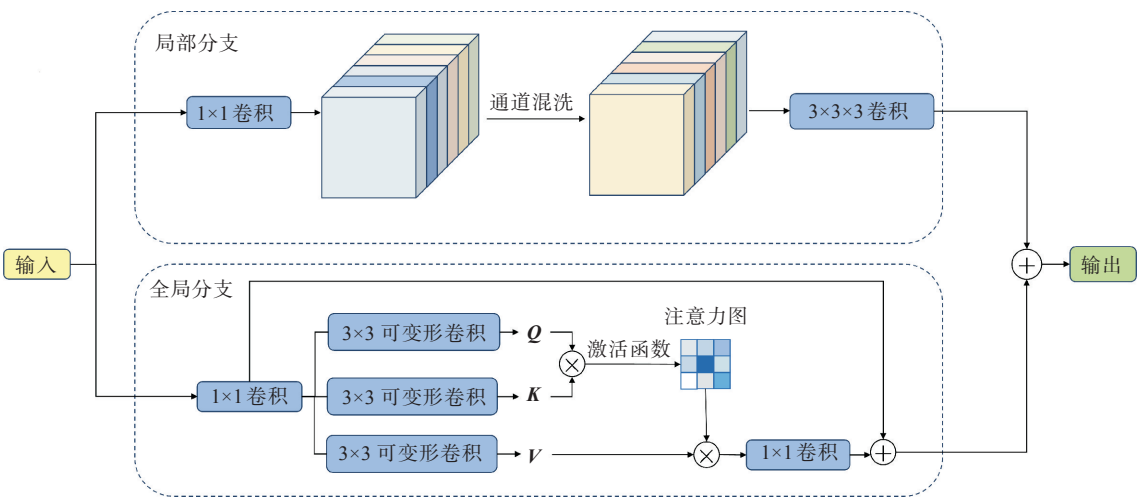

图 4　CAFM 结构示意图

Fig. 4　CAFM structure

检测头缺乏上下文融合机制，其逐位置独立预测的方式限制了模型对全局语义信息的感知能力。针对上述问题，本研究引入基于注意力机制的DyHead[16]，通过动态特征加权提升多尺度目标检测性能。

DyHead以主干网络输出的三维特征张量为基础，通过尺度、空间和任务3个维度的注意力机制协同工作，实现多尺度特征的动态融合。尺度感知注意力通过对特征金字塔中不同尺度特征进行加权融合，增强模型对目标尺度变化的适应性。空间位置感知注意力通过跨尺度集成特征，利用可变形卷积捕获跨区域上下文信息，提升密集目标的定位精度。任务感知注意力使用超参数控制激活阈值，使模型能够根据不同下游任务的需求灵活选择相关特征。检测头能够在保持实时性的同时提升对水下模糊目标、遮挡目标的检测鲁棒性，Dyhead的计算表达式如下：

$$W(\boldsymbol{F}) = \pi_D\left(\pi_S\left(\pi_L(\boldsymbol{F})\cdot\boldsymbol{F}\right)\cdot\boldsymbol{F}\right)\cdot\boldsymbol{F} \tag{8}$$

式中：$\boldsymbol{F}$表示主干网络输出的三维张量，其维度分别为通道维度D、空间维度S和水平维度L，π_D、π_S、π_L分别为任务感知注意力函数、空间感知注意力函数和尺度感知注意力函数。

2 结果与讨论

2.1 数据集

本研究选择了两个公开的水下光学数据集：RUOD[17]与DUO[18]，用于训练与测试水下目标检测模型。RUOD作为综合性水下目标检测基准数据集，包含14 000张图像和74 903个标注对象，涵盖鱼、潜水员、海星、珊瑚、海龟、海胆、海参、扇贝、鱿鱼、水母等10种常见的水生类别。该数据集反映了真实水下场景的特点，包括色彩畸变、低对比度和动态光干扰等。DUO数据集基于全国水下机器人专业大赛多届挑战数据构建，经冗余数据清洗与标签校正后形成标准化数据集。DUO数据集中包含7 782张精细标注图像和74 515个标注对象，专门针对海胆、海参、扇贝、海星4类经济性水产物种进行多场景覆盖。

这两个数据集在样本规模、目标类别和场景复杂度上相互补充。实验中，原始数据集按6∶2∶2的比例随机划分为训练集、验证集和测试集。

2.2 实验环境及参数

本实验硬件设备采用的GPU为NVIDIA GeForce RTX 4060 Ti（16 GB），CPU为intel i5-12600。软件环境为Window10操作系统，采用Pytorch作为深度学习框架，配置环境为：Cuda11.8+Python3.9+Pytorch2.0.0。所有实验都在相同的实验环境下执行，训练时采用随机梯度下降法进行优化，输入图片尺寸为640×640像素，批量大小为16，设置200个训练周期，设置初始学习率为0.01，动量为0.937。

2.3 评价指标

为了全面评估改进后模型的性能，从检测精度和速度两个方面进行分析，采用精确率、召回率、平均精度均值（mean average precision，mAP）、浮点运算次数（giga floating-point operations per second，GFLOPs）和参数量（parameters，Params）5个指标以准确、客观地评价模型的性能。其中，mAP表示所有类别的平均精度，值越高则模型性能越佳，目标识别精度越高；GFLOPs衡量浮点运算量，反映了模型的计算复杂度；Params表示模型中的参数总数，用于评估模型的规模。

2.4 对比实验

2.4.1 主干网络横向对比 对YOLOv11n的主干网络进行横向对比实验，以验证主干网络性能，实验结果如表1所示。EfficientFormerV2在检测精度上表现最优，交并比阈值为50%时的mAP（mAP@50）和mAP@50−95明显高于其他网络。相比之下，FasterNet性能次之，MobileNetV4和EfficientViT的表现相对较差，表明其模型结构设计在水下目标检测任务中的适应性不足。在计算效率方面，EfficientFormerV2同样表现出色，其参数量仅为4.84×10^6，GFLOPs为10.2，在保持较高精

表1 主干网络性能横向对比

Table 1 Horizontal comparison of backbone networks' performances

主干网络	mAP@50		mAP@50−95		参数量 / 10^6	GFLOPs
	RUOD	DUO	RUOD	DUO		
MobilenetV4	0.773	0.743	0.524	0.533	5.43	21.0
EfficientViT	0.827	0.792	0.580	0.587	3.14	7.9
FasterNet	0.842	0.821	0.594	0.623	3.90	9.2
EfficientFormerV2	0.846	0.824	0.603	0.625	4.84	10.2

度的同时具备较低的运算开销，适合计算资源受限的水下检测场景部署。综合检测精度和计算效率分析，EfficientFormerV2被证明是主干网络改进的最佳选择，能够有效平衡性能和资源消耗，适用于高效目标检测任务。

2.4.2　检测头横向对比　针对YOLOv11n的检测头进行了对比实验，将DyHead检测头与当前主流的检测头改进方案进行对比，实验结果如表2所示。DyHead在两个数据集上mAP@50均显著优于其他检测头，这表明其在特征提取和融合能力方面具有明显优势。虽然DyHead的参数量和计算量相比其他检测头略有增加，但其在检测精度上的提升有效弥补了额外的计算开销。相比之下，MultiSEAMHead和SEAMHead表现中等，而计算复杂度最低的EfficientHead检测精度不足。综合多方面性能评估，DyHead在水下目标检测任务中具有显著优势。

2.4.3　不同模型对比　为验证EAD-YOLOv11n模型的有效性，实验选取了两阶段检测器Faster R-CNN、单阶段检测器SSD以及轻量化的YOLO系列作为基准，并在RUOD和DUO数据集上进行了评估。

表3结果显示，EAD-YOLOv11n在保持较低计算量和参数量的同时，检测精度显著提升。与Faster R-CNN相比，EAD-YOLOv11n在RUOD数据集上的mAP@50和mAP@50–95分别提高了10.0%和13.7%。与SSD相比，EAD-YOLOv11n在DUO数据集上的mAP@50和mAP@50–95分别提升了14.1%和20.7%，同时大幅降低了计算开销。与轻量化模型相比，EAD-YOLOv11n在精度和效率上表现出均衡优势。虽然YOLOv8n和YOLOv10n的参数量和运算量较低，但EAD-YOLOv11n在RUOD和DUO数据集上的mAP@50分别达到了0.852和0.830。此外，EAD-YOLOv11n的参数量和运算量优于YOLOv7-tiny，进一步证明其在高运算效率下显著提升了检测精度。与YOLOv11n相比，EAD-YOLOv11n在RUOD数据集上的mAP@50和mAP@50–95分别提升1.3%和1.7%，在DUO数据集上提升1.8%和2.8%，且仅增加3.4×10^6参数量。实验表明，EAD-YOLOv11n在满足实时性需求的同时，实现了更高的检测精度。

2.5　消融实验

通过消融实验验证了EAD-YOLOv11n模型中各项改进的有效性及其组合对性能的影响，结果如表4所示。EfficientFormerV2主干网络显著提升了模型的准确率和召回率，但增加了一定的计算量和参数量。ACFM模块通过改进特征融合网络，抑制背景干扰，提高了定位精度，在RUOD和DUO数据集上精确率分别提升了0.5%和1.8%。DyHead模块利用注意力机制降低了漏检率，各项指标均有提升。实验结果表明，在计算代价可控的前提下，EAD-YOLOv11n模型在RUOD和DUO数据集上的准确率、召回率和平均精度均

表2　检测头性能横向对比

Table 2　Horizontal comparison of the detection heads' performances

检测头	mAP@50		mAP@50–95		参数量 / 10^6	GFLOPs
	RUOD	DUO	RUOD	DUO		
MultiSEAMHead	0.833	0.805	0.585	0.601	4.60	6.2
SEAMHead	0.835	0.809	0.588	0.605	2.49	5.8
EfficientHead	0.837	0.807	0.592	0.601	2.32	5.1
Dyhead	0.841	0.814	0.595	0.610	3.10	7.4

表3　改进模型与其他模型对比

Table 3　Comparison of the enhanced model with other models

模型	mAP@50		mAP@50–95		参数量 / 10^6	GFLOPs
	RUOD	DUO	RUOD	DUO		
Faster R-CNN	0.752	0.735	0.473	0.481	123.60	213.3
SSD	0.723	0.689	0.451	0.427	71.60	87.5
YOLOv7-tiny	0.836	0.817	0.548	0.584	6.02	13.2
YOLOv8n	0.836	0.810	0.588	0.610	3.01	8.1
YOLOv10n	0.827	0.801	0.574	0.596	2.27	6.5
YOLOv11n	0.839	0.812	0.593	0.606	2.59	6.5
EAD-YOLOv11n	0.852	0.830	0.610	0.634	5.99	12.8

表4 消融实验结果对比

Table 4 Comparison of ablation experimental results

模型	精确率		召回率		mAP@50		mAP@50−95	
	RUOD	DUO	RUOD	DUO	RUOD	DUO	RUOD	DUO
YOLOv11n	0.848	0.819	0.736	0.727	0.839	0.812	0.593	0.606
YOLOv11n+EfficientFormerV2	0.849	0.839	0.781	0.750	0.846	0.824	0.603	0.625
YOLOv11n+ACFM	0.853	0.837	0.766	0.723	0.843	0.810	0.599	0.608
YOLOv11n+DyHead	0.841	0.827	0.780	0.726	0.841	0.814	0.595	0.61
YOLOv11n+EfficientFormerV2+ACFM	0.848	0.817	0.784	0.755	0.846	0.826	0.605	0.631
YOLOv11n+EfficientFormerV2+ACFM+DyHead	0.848	0.831	0.789	0.754	0.852	0.830	0.610	0.634

达到最高水平。

2.6 可视化结果分析

本研究提出的改进模型在水下复杂场景中的检测效果较基线模型有显著提升，如图5所示。在目标密集分布场景中，改进后的模型能够清晰识别被海胆群遮挡的海参；在目标相互遮挡场景下，改进模型能准确区分海龟和周边的鱼群；在低可见度的模糊场景中，改进模型成功检测出沙土悬浮物遮挡下的贝壳轮廓。可视化结果表明，该模型能有效解决水下目标检测中遮挡误判问题和小目标漏检问题，为复杂水下环境下的生物监测提供了可靠的技术方案。

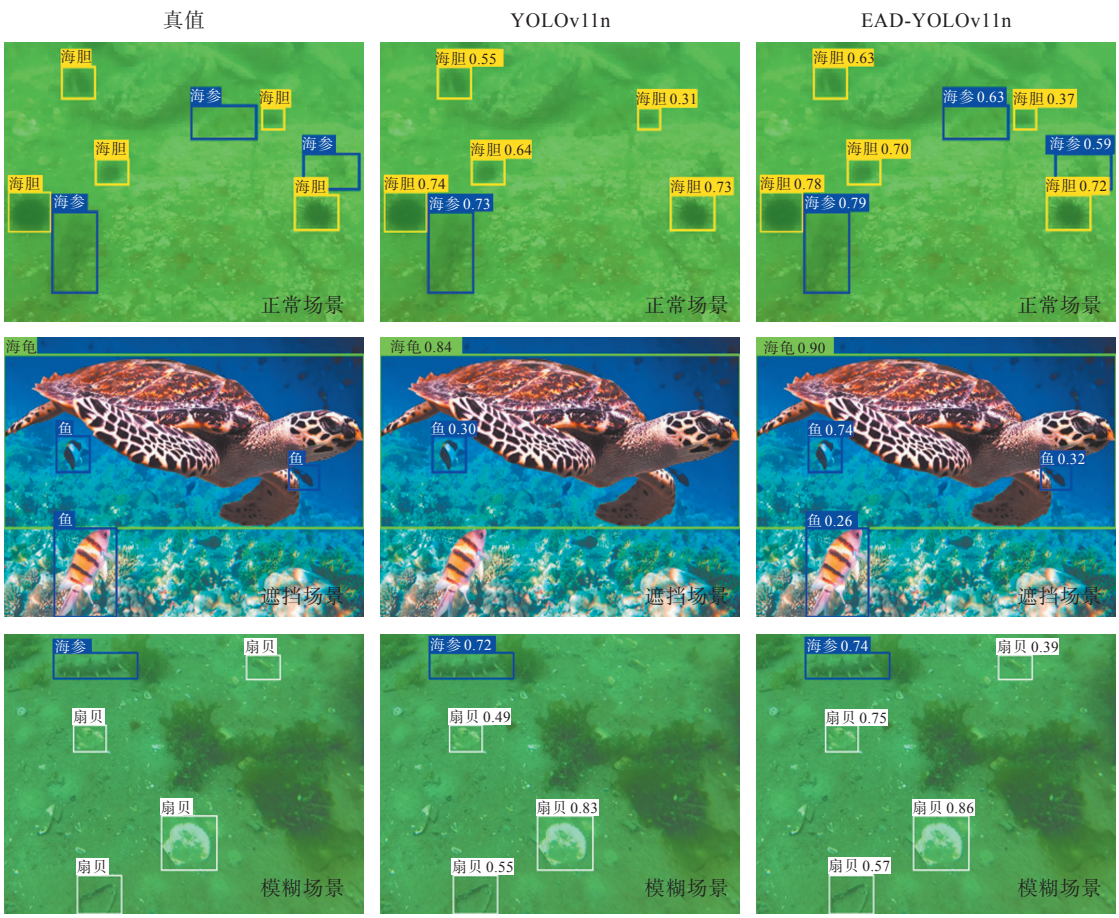

图5 模型改进前后检测效果对比

Fig. 5 Comparison of detection performances before and after model improvement

3 结 论

复杂水下环境中的目标检测任务面临多重挑战，如光照不均造成的色彩失真、目标与背景对比度不足，以及目标形态和尺寸的显著差异等。针对这些问题，本文提出一种基于YOLOv11n的复杂环境下水下目标检测模型EAD-YOLOv11n。通过引入EfficientFormerV2轻量化主干网络，提高对局部细节和全局语义信息的捕捉能力。同时，通过优化特征融合网络和替换检测头，进一步提

升了对多尺度水下生物目标的检测精度。改进后的模型有效解决水下模糊目标特征退化、跨通道信息整合不足以及多尺度目标识别精度低等问题。在RUOD和DUO数据集上的实验表明，该模型的mAP@50分别达到0.852和0.830，较基准模型提升1.3%和1.8%，且参数量仅增加3.4×10^6。该模型的性能优于现有的主流模型，同时保持了相对较低的计算复杂度。未来，将重点优化模型架构，进一步压缩计算规模，以实现边缘计算设备的高效部署。

参考文献

[1] 王德兴，何勇，袁红春. 基于YOLOv8-BAN模型的水下生物目标检测方法[J]. 江苏农业学报，2025，41(1)：101-111.

[2] 孙艺倩. 基于深度学习的水下垃圾检测方法研究[D]. 吉林：东北电力大学，2023.

[3] 张有波. 基于视觉的水下考古机器人实时目标检测与识别[D]. 上海：上海海洋大学，2021.

[4] XU S B，ZHANG M H，SONG W，et al. A systematic review and analysis of deep learning-based underwater object detection [J]. Neurocomputing，2023，527：204-232.

[5] 黄瑜豪，曾祥进，冯崧. 面向边缘设备的轻量级OpenPose姿态检测模型研究[J]. 武汉工程大学学报，2024，46(4)：424-430.

[6] REN S Q，HE K M，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，2017，39(6)：1137-1149.

[7] HE K M，GKIOXARI G，DOLLÁR P，et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway，NJ：IEEE，2017：2980-2988.

[8] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：unified，real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway，NJ：IEEE，2016：779-788.

[9] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot multibox detector[C]//Computer Vision-ECCV 2016. Berlin：Springer，2016：21-37.

[10] 陈宇梁，董绍江，孙世政，等. 改进YOLOv5s的弱光水下生物目标检测算法[J]. 北京航空航天大学学报，2024，50(2)：499-507.

[11] 辛世澳，葛海波，袁昊，等. 改进YOLOv7的轻量化水下目标检测算法[J]. 计算机工程与应用，2024，60(3)：88-99.

[12] 李培坤，李锋，葛忠显，等. 基于改进YOLOv8n的水下目标检测算法[J]. 电子测量技术，2025，48(3)：172-179.

[13] LI Y Y，HU J，WEN Y，et al. Rethinking vision Transformers for MobileNet size and speed [C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway，NJ：IEEE，2023：16843-16854.

[14] CHEN Z X，HE Z W，LU Z M. DEA-Net：single image dehazing based on detail-enhanced convolution and content-guided attention [J]. IEEE Transactions on Image Processing，2024，33：1002-1015.

[15] HU S，GAO F，ZHOU X W，et al. Hybrid convolutional and attention network for hyperspectral image denoising [J]. IEEE Geoscience and Remote Sensing Letters，2024，21：5504005.

[16] DAI X Y，CHEN Y P，XIAO B，et al. Dynamic head：unifying object detection heads with attentions [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2021：7369-7378.

[17] FU C P，LIU R S，FAN X，et al. Rethinking general underwater object detection：datasets，challenges，and solutions[J]. Neurocomputing，2023，517：243-256.

[18] LIU C W，LI H J，WANG S C，et al. A dataset and benchmark of underwater object detection for robot picking [C]//2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Piscataway，NJ：IEEE，2021：1-6.

本文编辑：霍淑利