

文章编号 1004-924X(2026)11-1762-14

## 面向复杂水环境的轻量化多尺度漂浮物检测

杨天麒, 王 勇\*

(中国地质大学(武汉)机械与电子信息学院, 湖北 武汉 430074)

**摘要:**针对复杂水环境中漂浮物检测易出现误检、漏检及多尺度感知不足的问题,提出一种基于YOLOv8n的轻量化高效检测方法。在主干网络及颈部深层设计融合重参数化卷积(RepConv)的多尺度自适应上下文聚合模块MAC-ELAN(Multi-scale Adaptive Context Enhanced Layer Aggregation Network),通过跨层特征融合与上下文增强提升复杂背景下的特征判别能力,并利用结构重参数化兼顾检测精度与计算效率;同时,引入P2与P6检测头增强模型对极小与大尺度目标的感知能力。在颈部浅层设计轻量化特征融合模块C2f-FA,通过部分卷积(PCConv)与EMA注意力机制重构C2f结构,并结合Wise-IoUv3损失提升目标定位精度。实验结果表明,在自建数据集上,改进模型仅增加1.13M参数,使mAP@50与mAP@50:95分别提升4.2%和3.9%。在Flow-Img与IWHR-Floater\_V1数据集上的跨域实验进一步验证了该方法在复杂水面场景中的有效性,在保持较低计算复杂度的同时实现了检测精度与模型轻量化之间的良好平衡。

**关键词:**漂浮物检测;复杂水环境;重参数化卷积;特征融合;轻量化网络

中图分类号:TP391 文献标识码:A

doi:10.37188/OPE.20263411.1762

CSTR:32169.14.OPE.20263411.1762

## Lightweight multi-scale floating detection for complex water environments

YANG Tianqi, WANG Yong\*

(School of Mechanical and Electronic Information, China University of Geosciences,  
Wuhan 430074, China)

\* Corresponding author, E-mail:wy112708@163.com

**Abstract:** Floating object detection in complex water environments is challenged by false positives, missed detections, and limited multi-scale perception. A lightweight and efficient method based on YOLOv8n is proposed. A Multi-scale Adaptive Context-Enhanced Layer Aggregation Network (MAC-ELAN), integrating reparameterized convolution (RepConv), is introduced into the backbone and the deeper layers of the neck. Through cross-layer feature fusion and contextual enhancement, feature discrimination in complex backgrounds is strengthened, while structural reparameterization balances detection accuracy and computational efficiency. In addition, P2 and P6 detection heads are incorporated to improve sensitivity to extremely small and large-scale targets. In the shallow layers of the neck, a lightweight feature fusion module, C2f-FA, is designed by reconstructing the C2f structure with Partial Convolution (PCConv) and the EMA attention mechanism. The Wise-IoUv3 loss is further adopted to enhance localiza-

收稿日期:2026-03-18;修订日期:2026-04-01.

基金项目:国家自然科学基金资助项目(No. 61973283)

tion accuracy. Experimental results on a self-constructed dataset demonstrate that, compared with the baseline, the proposed model improves mAP@50 and mAP@50:95 by 4.2% and 3.9%, respectively, with only 1.13M additional parameters. Cross-domain evaluations on the Flow-Img and IWHR-Floater\_V1 datasets further verify the effectiveness of the proposed approach in complex water-surface scenarios, achieving a superior balance between detection accuracy and model lightweightness while maintaining low computational complexity.

**Key words:** floater detection; complex water environment; reparameterized convolution; feature fusion; lightweight network

## 1 引言

水面漂浮垃圾污染<sup>[1]</sup>已成为威胁水生生态系统健康与公共安全的重要环境问题,构建高效、可靠的漂浮物自动监测系统具有重要意义。早期研究多依赖传统图像处理与机器学习方法,如基于背景建模的前景提取和视觉显著性分析方法。Zuo等<sup>[2]</sup>基于高斯混合模型进行背景建模,并结合颜色与直方图特征实现漂浮物检测。Sun等<sup>[3]</sup>融合空间纹理与频域显著性信息,通过低秩分解抑制水面纹理干扰以提升检测效果。然而,此类方法依赖人工特征与先验知识,对水面倒影及目标外观变化适应性较差,且在复杂背景与多尺度目标条件下稳定性和泛化能力有限,难以满足实际应用需求。

相比传统方法,基于深度学习的目标检测器能够通过端到端训练自动学习具有强判别力的高层特征,从而在尺度与形态多样、背景干扰复杂的水面环境中实现更加稳定的漂浮物检测。现有深度学习目标检测方法主要分为双阶段与单阶段两类。其中,双阶段方法以R-CNN系列为代表,其核心思想是先生成候选区域,再对候选目标进行精细分类与回归。Li等<sup>[4]</sup>基于Faster R-CNN提出了金字塔锚生成策略与高分辨率类别判别图,使检测器能够根据目标尺寸自适应生成更匹配的锚框,在抑制复杂背景干扰的同时显著增强了小尺度漂浮物的判别能力。然而,双阶段方法通常依赖复杂的候选区域生成与特征对齐过程,计算与存储开销较大,难以满足轻量化与实时部署需求。

单阶段检测方法以SSD<sup>[5]</sup>和YOLO<sup>[6]</sup>为代表,其通过密集预测实现端到端目标检测,具有推理速度快、部署灵活等优势,因而在水面漂浮物检测领域得到广泛应用。Chen等<sup>[7]</sup>将改进后的SSD与自适应滤波相结合,实现了检测与跟踪的一体

化设计,在检测精度与目标轨迹连续性方面均取得提升。YOLO系列因其优越的性能和工程友好性,已成为目标检测领域应用最广泛的框架,YOLOv5是其中的典型代表。相关改进工作主要通过优化主干网络、引入注意力机制、设计新型特征融合模块以及拓展检测头等方式来提升性能<sup>[8-11]</sup>。然而,此类方法往往以增加网络复杂度为代价,在提升精度的同时引入额外的计算开销,使模型在资源受限的边缘设备上部署受限。为此,部分研究开始转向轻量化漂浮物检测算法的探索,主要通过轻量化模块替换、模型剪枝以及改进损失函数与样本采样策略等手段降低模型复杂度<sup>[12-16]</sup>。尽管上述方法在一定程度上缓解了效率与精度之间的矛盾,但在兼顾多尺度特征表达、复杂水面背景抑制与模型轻量化方面仍显不足。尤其在目标尺度差异显著、倒影等干扰强烈的真实水面场景中,检测性能有待进一步提升。

针对上述技术难点,本文基于YOLOv8n提出了一种高效的水面漂浮物检测方法。该方法通过性能增强与效率优化的协同设计实现精度与轻量化的平衡。在性能方面,引入多尺度自适应上下文增强聚合模块(MAC-ELAN),结合重参数化卷积与分层注意力机制以强化特征提取以及抗干扰能力,并通过扩展特征金字塔结构,增加检测头以构建多粒度检测体系;同时引入WIoUv3损失函数<sup>[15]</sup>,通过动态样本聚焦提升目标定位的稳定性,使得模型对于水环境更具有针对性。在效率方面,设计轻量化C2f-FA模块,利用部分卷积与EMA<sup>[17]</sup>注意力机制替换颈部冗余结构,在保持特征表达能力的同时有效降低计算开销。实验结果表明,该方法在保证模型轻量化的前提下,显著提升了复杂水环境中多尺度漂浮物的检测精度与泛化能力。

## 2 网络设计

近年来,目标检测领域不断涌现出新的检测框架,如 YOLO11 与 YOLO12 等模型在结构设计上进一步强化了轻量化特性,在参数规模与计算复杂度方面具有一定优势。然而,现有研究与工程实践表明,在中小规模数据集或特定场景任务中,这些新模型在检测精度与稳定性方面并不一定优于经典模型。相比之下, YOLOv8 系列在检测精度、训练稳定性以及工程部署成熟度方面表现更加均衡,其中 YOLOv8n 作为该系列中最轻量化的版本,在保持较低参数数量和计算复杂度

的同时,仍能够提供较为可靠的检测性能,因此,在实际工程应用中得到了更为广泛的使用。然而,在复杂水面环境中,由于水体倒影以及目标尺度变化显著等因素的影响,即使是轻量化的 YOLOv8n 在实际检测过程中仍可能出现漏检与误检问题。因此,有必要结合水面场景的成像特性与漂浮物目标分布特点,对 YOLOv8n 的网络结构与特征表达机制进行针对性优化,以在保持模型轻量化优势的同时进一步提升其在复杂水环境中的检测稳定性与可靠性。

### 2.1 网络结构

本文提出的改进方法整体如图 1 所示。首

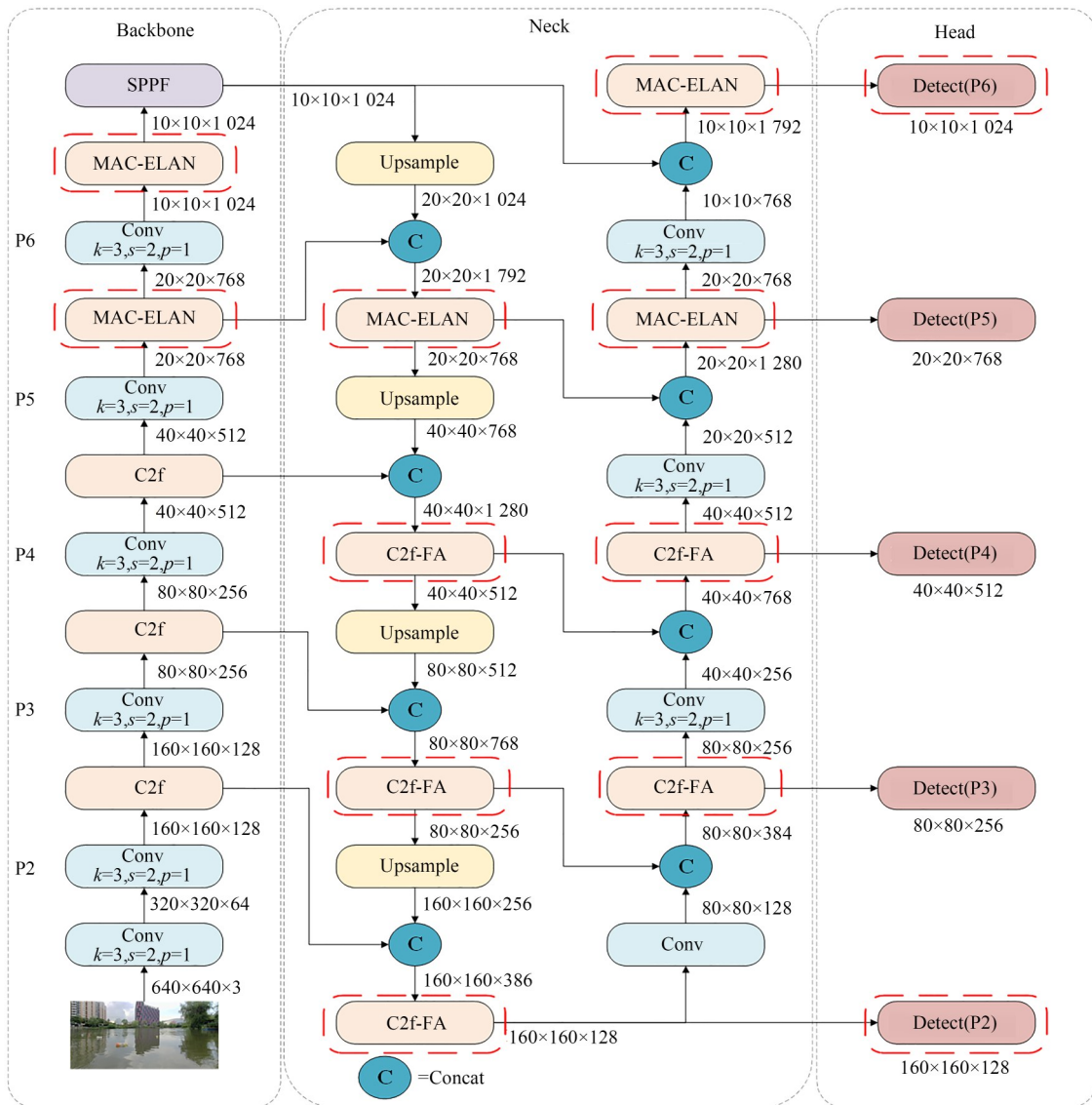


图 1 改进的漂浮物检测模型结构

Fig. 1 Architecture of improved floating detection model

先,设计MAC-ELAN模块并将其引入主干网络的深层,该模块通过聚合不同感受野的特征,使网络能够同时捕获目标的局部细节与其宏观的空间分布模式,显著提升对多尺度目标的表征能力。模块采用的重参数化卷积(RepConv)<sup>[18]</sup>不仅强化了模型特征提取能力,还优化了梯度传播路径,有效缓解了深层网络中的梯度退化问题,其集成的通道注意力机制可自适应抑制水面倒影等背景干扰。其次,在颈部网络进行差异化模块设计,在特征金字塔顶端采用MAC-ELAN模块,利用其强大的上下文建模能力为经过多次下采样的深层特征补充全局语义信息,确保输出特征具备更强的判别力。在特征金字塔中下层引入C2f-FA模块,旨在以轻量化的方式传递与初步筛选多尺度特征,从而保障高分辨率细节向高层网络的流动。这种组合在提升性能的同时合理地控制了模型的参数量和计算开销。然后,对于检测头部分,将检测头扩展为包含P2与P6层的多尺度输出结构,使模型能够同时适应小尺寸漂浮物与大范围污染区域等多尺度目标的检测需求。最后,为了优化模型对受到复杂背景干扰的困难正样本的学习能力,同时增强训练稳定性,采用WIoUv3来替代CIoU作为损失函数。上述改进使模型能够在复杂水面环境下保持较高的检测精度与运行效率,较好地平衡了性能与轻量化需求。

## 2.2 MAC-ELAN 模块

在保证计算效率的前提下提升多尺度漂浮物的特征表征能力,本文基于ELAN设计了MAC-ELAN模块。该模块通过层级化RepConv建模与跨尺度特征聚合扩展感受野,并结合通道注意力对特征进行自适应重标定,从而增强关键目标区域的表达能力。在较低计算开销下,MAC-ELAN有效提升了复杂水面环境中多尺度目标的表征能力,其结构如图2所示。

具体而言,设输入特征图为 $X \in \mathbb{R}^{C \times H \times W}$ 。首先,通过 $1 \times 1$ 卷积层进行初步通道映射和初步特征整合,随后在通道维度上将特征划分为两条并行路径。其中,一条主路径通过多个串联的RepConv模块进行逐层特征建模,通过不断扩展感受野逐步提取更高层语义信息,并保留各层输出以构建多尺度特征集合;另一条路径则直接保

留原始空间特征,以维持高分辨率结构信息并减少细节损失。最终,两条路径产生的特征在通道维度进行拼接,实现由局部纹理信息到高层语义信息的连续特征聚合。该多尺度聚合方式能够在保持结构轻量化的同时显著增强网络的上下文建模能力,使模型能够更加充分地捕获漂浮物目标的形态特征与空间分布信息。

为了进一步提升特征表达质量,模块引入通道注意力机制对融合特征进行自适应重标定。首先通过全局平均池化获取通道级上下文描述,并利用两层 $1 \times 1$ 卷积生成通道注意力权重,其计算过程为:

$$s = \sigma \left( \text{Conv}_2 \left( \delta \left( \text{Conv}_1 \left( \text{GAP} \left( F_{\text{cat}} \right) \right) \right) \right) \right), \quad (1)$$

其中: $\delta(\cdot)$ 表示SiLU激活函数, $\sigma(\cdot)$ 表示Sigmoid函数, $\text{GAP}(\cdot)$ 为全局平均池化, $F_{\text{cat}}$ 则为多尺度聚合后的特征。该注意力机制使网络能够自适应强化与目标相关的特征响应,同时抑制复杂背景区域带来的干扰,从而提升特征判别能力。最终,通过 $1 \times 1$ 卷积层完成通道压缩并输出融合特征。

RepConv作为MAC-ELAN核心特征提取模块,在训练阶段采用多分支卷积结构,包括 $3 \times 3$ 卷积分支、 $1 \times 1$ 卷积分支以及恒等映射分支。其输出形式为:

$$y = f_{3 \times 3}(x) + f_{1 \times 1}(x) + f_{\text{id}}(x), \quad (2)$$

其中: $x$ 为输入特征, $y$ 为输出。恒等映射分支由BN来近似实现。该结构通过增加特征传播路径与隐式特征集成机制,有效增强了特征表达能力和梯度传播效率,从而提升网络在复杂场景下的特征学习能力。在推理阶段,其可等效重参数化为单一标准卷积,从而保证较高的计算效率。在推理阶段,通过结构重参数化将多分支卷积简化成单个 $3 \times 3$ 卷积,其等效的卷积核可表示为:

$$W_{\text{eq}} = W_{3 \times 3} + \text{pad}(W_{1 \times 1}) + W_{\text{id}}, \quad (3)$$

其中: $W_{\text{eq}}$ 为重参数化后得到的等效卷积核, $W_{3 \times 3}$ 与 $W_{1 \times 1}$ 分别为训练阶段 $3 \times 3$ 卷积分支和 $1 \times 1$ 卷积分支得到的卷积核权重, $W_{\text{id}}$ 则为恒等映射分支的卷积核权重。通过上述参数融合操作,训练阶段的多分支结构可以在推理阶段完全等效为单个卷积层,从而在保持特征表达能力的同时显著降低推理计算开销。

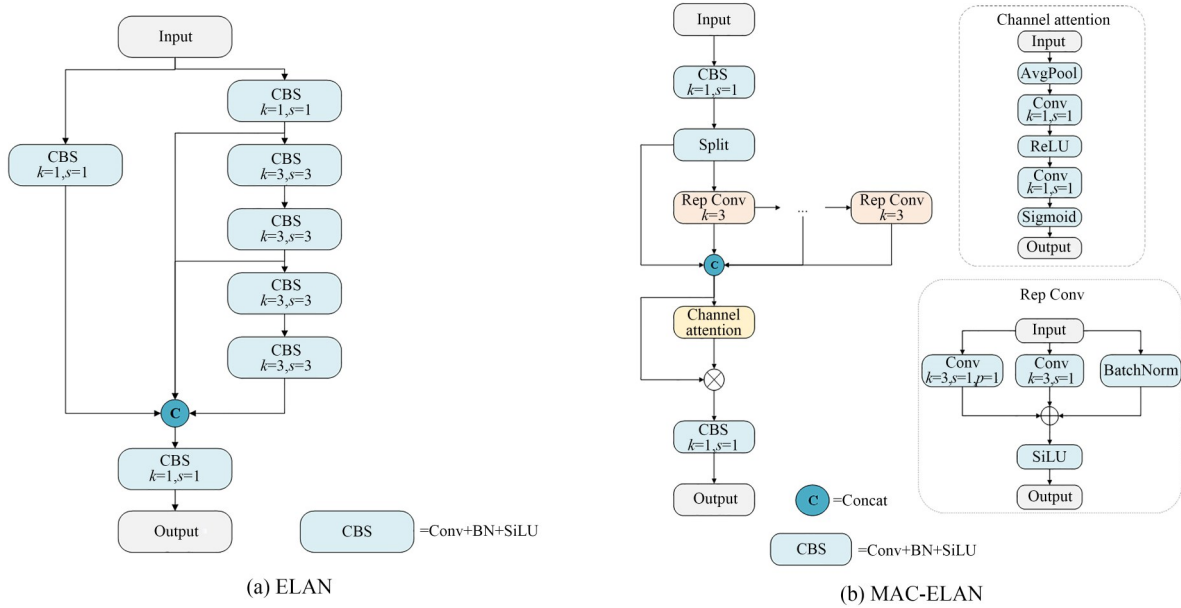
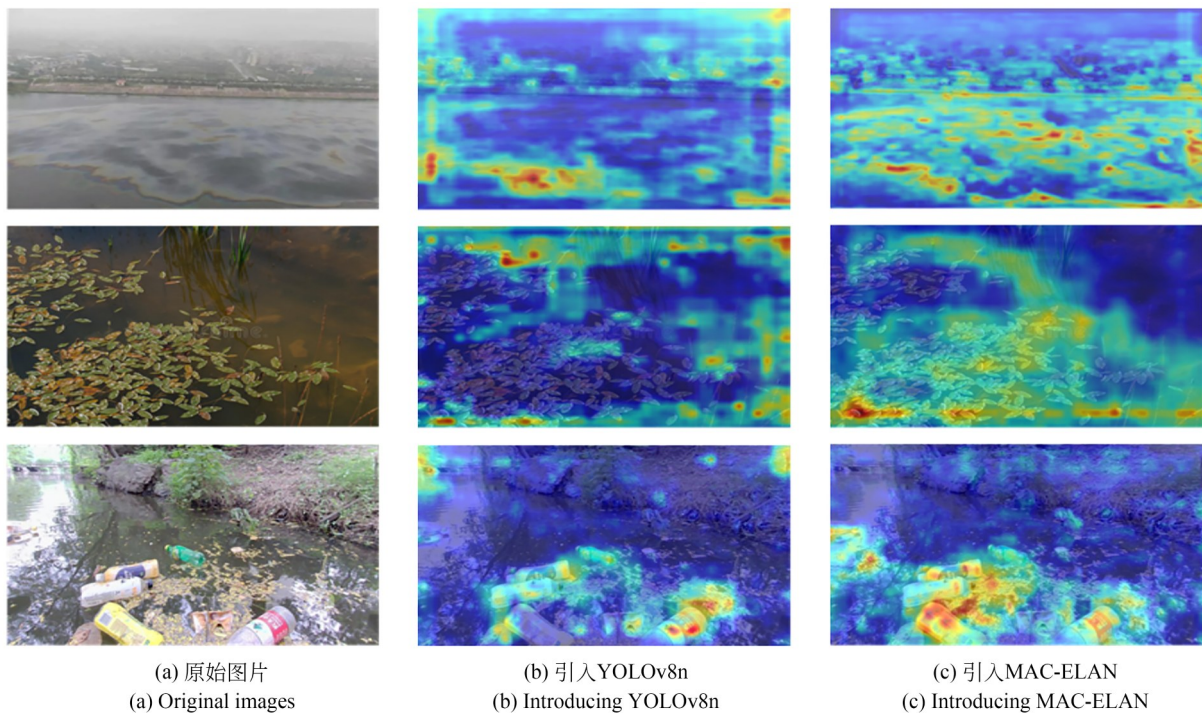


图 2 ELAN 和 MAC-ELAN 的结构  
Fig. 2 Architecture of ELAN and MAC-ELAN

图 3 对比了引入 MAC-ELAN 模块前后网络的特征响应热力图。从图中可以看出,在面对大面积目标时,基准模型的注意区域较为分散,往往仅集中于目标的局部区域,甚至出现对非目标区域的响应。而在引入 MAC-ELAN 模块后,模

型能够更加准确地聚焦于目标主体区域,整体特征响应更加完整且集中。该结果表明,该模块能够有效增强网络的特征表征能力,并强化目标边缘与关键区域的响应,从而显著提升模型在水面场景中应对目标尺度变化及复杂背景干扰的能



(a) 原始图片 (a) Original images  
(b) 引入YOLOv8n (b) Introducing YOLOv8n  
(c) 引入MAC-ELAN (c) Introducing MAC-ELAN

图 3 引入 MAC-ELAN 前后的检测结果热力图对比  
Fig. 3 Comparison of detection result heatmaps before and after introduction of MAC-ELAN

力。在引入该模块后,网络参数量由 3.01 M 降至 2.57 M,在提升特征表达能力的同时实现了模型的进一步轻量化。

### 2.3 检测头

针对水面漂浮物检测任务中极端尺度分布,例如微小塑料瓶与大面积油污共存的技术难点,本文对 YOLOv8 的检测头结构进行了针对性扩展,将原本的三检测头增设为五检测头。原结构在对应极端尺度目标时存在局限,深层特征图难以保留小目标的细节纹理,而浅层特征又缺乏对大目标的全局感知能力,导致小目标漏检与大目标定位不准。新增的 P2 检测头位于特征金字塔顶端,其较高的空间分辨率有助于保留细小目标的纹理与边缘特征,有效提升了模型对像素级漂浮物的敏感度;P6 检测头则依托更大的感受野,能够整合图像区域的宏观语义信息,从而更准确地捕捉大面积污染区域的整体轮廓与空间分布。这两个检测头的引入构建了多粒度协同的检测机制,使其能够同时覆盖从局部细节到全局结构的不同尺度目标表达需求,在复杂水面环境下显著增强了目标检测的完整性与稳定性。

### 2.4 C2f-FA 模块

为平衡因引入 P2 与 P6 检测头而增加的计算开销,本文在颈部网络中提出一种轻量级特征融合模块 C2f-FA,其结构如图 4 所示。该模块以融合 FasterNet 部分卷积<sup>[19]</sup>与 EMA 注意力的 Fast-

erBlockEMA 单元替换原 Bottleneck 结构为核心改进,在保证特征表达能力的同时显著降低了参数量与计算复杂度。该模块首先通过  $1 \times 1$  卷积层对通道数进行调整,随后在通道维度上将特征均分为两条分支。其中,一条分支依次通过 3 个串联的 FasterBlockEMA 模块以进行深层特征建模,该模块内部采用部分卷积策略,默认设置只对  $1/4$  通道进行卷积运算,从而显著降低冗余运算,同时引入 EMA 注意力机制,默认将通道分为 8 组,并通过自适应平均池化实现空间与通道信息的联合建模,从而增强对关键区域的响应能力并抑制水面反射等噪声干扰。需要特别指出的是,与传统固定尺寸池化核不同,本文所采用的自适应平均池化能够根据输入特征图尺寸动态调整感受野范围,在空间维度上实现全局上下文信息聚合,因此不存在显式的池化核大小参数。这种设计不仅避免了对固定感受野的依赖,同时更适用于水面场景中尺度变化较大及反射干扰显著的复杂环境,有助于增强模型对关键区域的响应能力并抑制噪声干扰。

另一条分支则直接保留原始特征,以保证信息传递的完整性。两条分支的输出在通道维度上进行拼接,并通过  $1 \times 1$  卷积层完成特征融合。该结构在继承 C2f 模块高效梯度传播特性的基础上,通过减少冗余卷积计算实现了更优的计算效率,达到了更好的平衡。

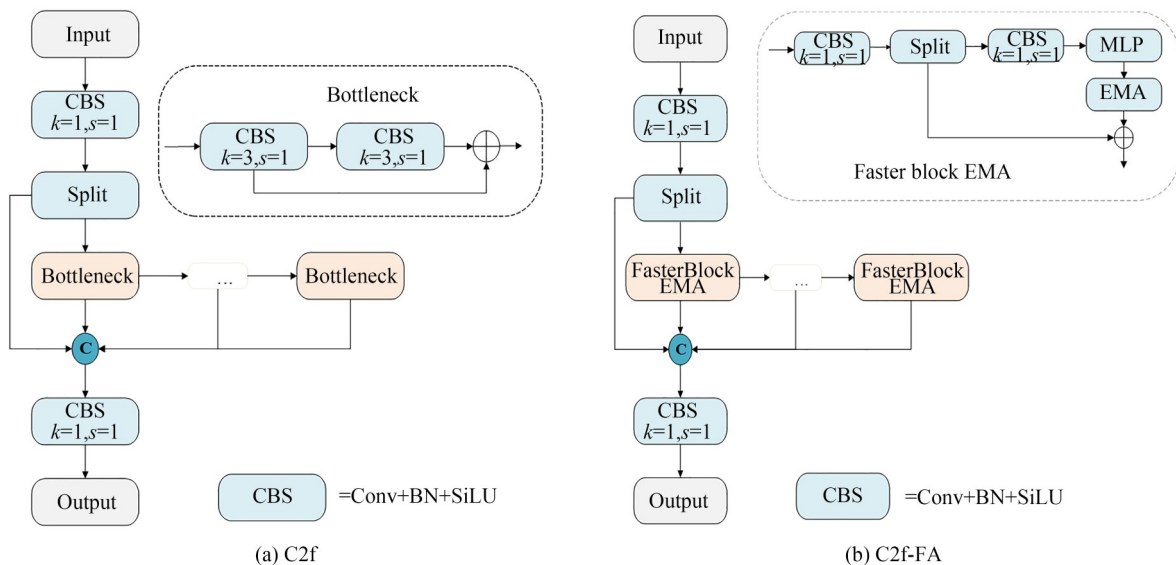


图 4 C2f 和 C2f-FA 的结构  
Fig. 4 Architecture of C2f and C2f-FA

作为模块的核心计算单元, FasterBlockEMA 采用 PConv 与注意力机制协同的轻量化结构。相较于常见的轻量化卷积网络, 尽管深度可分离卷积等方法能够在一定程度上减少参数量, 但其仍然存在较为频繁的内存访问开销, 从而限制了实际计算效率。相比之下, PConv 在降低参数规模的同时能够有效减少内存访问次数, 从而实现更高效的特征提取, 其结构如图 5 所示。具体而言, PConv 仅对输入特征图的一部分通道执行常规卷积运算, 而其余通道保持不变。对于连续或常规的内存访问, 仅需计算其中一个连续通道并将其视为整个特征图的代表, 从而显著减少计算与访存开销。设常规卷积的通道数为  $C$ , PConv 参与计算的通道数为  $c_p$ , 对于标准卷积, 其浮点运算量可表示为:

$$F = h \times w \times k^2 \times c_p^2, \quad (4)$$

其中:  $h$  与  $w$  分别表示特征图的高度和宽度,  $k$  表示卷积核的空间尺寸,  $F$  表示浮点运算次数 (FLOPs)。采用 PConv 后, 其计算复杂度近似为:

$$F' = h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p. \quad (5)$$

可以看出, PConv 在显著减少计算量的同时仍能够保留必要的特征表达能力。

在此基础上, FasterBlockEMA 进一步结合注意力机制以增强特征表达能力。输入特征首先通过  $1 \times 1$  卷积完成通道映射, 实现特征压缩与通道重组; 随后仅选取约  $1/4$  的通道参与深层特征建模, 该部分通道首先通过  $3 \times 3$  卷积提取局部空间信息, 然后通过两层 MLP 实现通道维度的非线性映射与特征重整。引入 EMA 注意力机制对特征进行自适应增强, 通过在水平方向与垂直方向进行自适应池化以提取方向感知的上下文信息, 并生成相应的特征权重, 从而强化关键区域响应。其余约  $3/4$  的通道通过恒等映射直接保留, 以减少不必要的卷积计算并避免冗余特征干扰。最终, 两部分特征通过残差连接进行融合并输出, 从而在保持信息完整性的同时提升特征表达能力。实验结果表明, 在基线模型中引入 C2f-FA 模块后, 模型参数量由 3.01 M 降至 2.94 M, FLOPs 由 8.1 G 降至 7.9 G。该模块在显著降低计算复杂度的同时, 仍然保持了良好的特征选择能力与上下文建模

能力, 从而在复杂水面环境中实现了推理效率与检测稳定性的有效平衡。

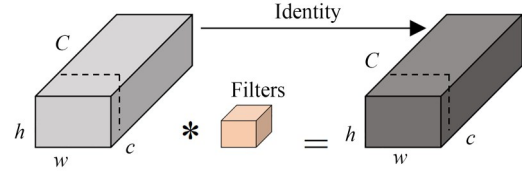


图 5 PConv 结构

Fig. 5 Structure of PConv

## 2.5 损失函数

在目标框回归过程中, CIoU 对不同类型样本采用均等优化策略。在复杂水面场景中, 该特性易受到大面积水面倒影等干扰区域的影响; 同时, 其尺度敏感的惩罚项倾向于对大目标施加更强约束, 从而削弱对小目标的优化效果, 导致小目标漏检率升高且定位精度不足。相比之下, WIoUv3 通过引入动态非单调聚焦机制, 使训练过程更加关注困难样本, 并在不同质量样本之间自适应地平衡损失贡献, 因此, 更适用于复杂水面环境下的多尺度漂浮物检测任务。WIoUv3 的计算公式为:

$$\mathcal{L}_{\text{WIoUv3}} = r \times \mathcal{L}_{\text{WIoUv1}}, \quad (6)$$

其中:  $\mathcal{L}_{\text{WIoUv1}}$  为距离惩罚项,  $r$  为非单调聚焦因子, 距离惩罚项定义为:

$$\mathcal{L}_{\text{WIoUv1}} = R_{\text{WIoU}} \times (1 - \text{IoU}), \quad (7)$$

其中:  $R_{\text{WIoU}}$  表示基于预测框与真实框中心点归一化距离的惩罚因子, 其计算形式为:

$$R_{\text{WIoU}} = \exp\left(\frac{(x - x_{\text{gt}})^2 + (y - y_{\text{gt}})^2}{(W_{\text{g}}^2 + H_{\text{g}}^2)^*}\right). \quad (8)$$

另一方面, 式(6)中非单调聚焦因子用于根据样本质量动态调整损失权重, 通过超参数  $\delta$  与  $\gamma$  调节对不同质量样本的关注程度, 其表达式为:

$$r = \frac{\beta}{\delta \gamma^{\beta - \delta}}. \quad (9)$$

$\beta$  为反映锚框质量的离群度量, 其定义为:

$$\beta = \frac{\mathcal{L}_{\text{IoU}}^*}{\mathcal{L}_{\text{IoU}}} \in [0, +\infty), \quad (10)$$

其中:  $\mathcal{L}_{\text{IoU}}^*$  为  $\mathcal{L}_{\text{IoU}}$  的滑动平均值, 用于刻画历史损失水平, 并通过与当前损失的相对关系间接反映样本难度, 从而为自适应加权提供依据。引入

WIoUv3后,模型能够对受水面倒影干扰而产生偏移的预测框施加更有效的惩罚,从而抑制此类干扰;同时,通过降低对简单样本的关注度并强化对困难样本的学习,缓解了目标尺度和外观差异带来的检测性能下降。WIoUv3具备更稳定的梯度特性与更低的计算复杂度,有助于模型的稳健收敛,并更好地满足边缘设备部署需求。

### 3 实验与分析

#### 3.1 数据集

本文所用的自制数据集基于网络公开资源与人工实地拍摄的水面图像构建而成。为提升数据多样性并增强模型的泛化能力,对原始图像进行了噪声扰动、亮度变化以及旋转、缩放等几何增强处理,最终构建了一个包含 7 772 张图像的高质量水面漂浮物检测数据集。该数据集包含塑料瓶、塑料袋、油污、蓝藻、落叶、荷花、鱼和荷叶八类典型水面目标,其尺度涵盖从微小碎片到大面积漂浮物,且形态差异显著,可有效验证模型在复杂水面场景中对多尺度目标的检测能力。数据集按照 8:1:1 的比例划分为训练集、验证集和测试集。

为进一步验证所提模型的有效性及其跨场景泛化性能,本文在两个公开水面目标检测数据集上开展了对比实验。Flow-Img 数据集<sup>[20]</sup>包含 2 000 张城市河道场景图像,共标注 5 271 个瓶类目标;IWHR\_Floater\_V1 数据集<sup>[21]</sup>由 3 000 张监控场景图像组成,标注目标数量达 23 692 个,涵盖塑料瓶、泡沫板及藻类等典型漂浮物。通过多数据集的综合实验设置,能够系统评估模型在不

同环境条件、污染物类型及目标尺度分布下的检测性能。

#### 3.2 算法性能评估

所有实验均在统一软硬件环境下进行,操作系统为 Windows 10, CPU 为 Intel Core i5-13400F, GPU 为 NVIDIA GeForce RTX 4060, 基于 PyTorch 2.0 与 CUDA 12.2 加速, Python 版本为 3.9。输入图像尺寸为  $640 \times 640$ , Batch Size 设为 16, 训练轮数为 300, 采用 SGD 优化器, 初始学习率为 0.01, 动量为 0.937, 并在训练最后 10 个 Epoch 关闭 Mosaic 增强以提升收敛稳定性。损失函数的关键超参数设置为  $\gamma = 1.9, \delta = 3$ , IoU 均值通过指数滑动平均更新, 并采用非单调聚焦机制。模型性能通过 mAP@50, mAP@50:95、精确率  $P$  及召回率  $R$  进行评估, 实验均在此配置下进行, 以保证结果的公平性与可复现性。

表 1 给出了所提方法与基准模型及当前最新 YOLO 系列模型的对比结果。为了使实验对比更加充分, 本文还将 GhostNet<sup>[22]</sup> 和 ShuffleNet<sup>[23]</sup> 分别引入 YOLOv8n 模型作为轻量化模型的代表进行对比分析。实验结果表明, 与基线模型相比, 所提方法在各项评价指标上均取得了明显提升, 其中 mAP@50 提高 4.2%, mAP@50:95 提高 3.9%, 精确率和召回率分别提升 5.1% 与 2.1%。进一步与其他模型进行比较, 所提方法在 mAP@50 与 mAP@50:95 上分别平均领先 3.7% 与 3.3%, 精确率和召回率分别平均提高 5.2% 与 2.3%。相较于对比模型中性能最优的 YOLO11s, 本文方法在参数量不足其一半的情况下, 仍实现了更优的整体检测性能。

在推理速度方面, 各模型在 GPU 平台上的

表 1 自制数据集的测试结果

Tab. 1 Test results on self-collected dataset

Model	mAP@50/%	mAP@50:95/%	$P$ /%	$R$ /%	Params/ M	FLOPs/ G	FPS/ (frame·s <sup>-1</sup> )
YOLOv8n	59.0	35.4	69.1	55.1	3.01	8.1	158
YOLOv8s	58.7	35.5	66.7	55.8	11.1	28.5	142
YOLO11s	60.7	38.1	70.7	55.4	9.43	21.6	148
YOLO12s	60.1	37.6	69.8	54.7	9.25	21.5	144
YOLOv8n+GhostNet	59.5	34.8	69.7	53.7	1.72	5.0	135
YOLOv8n+ShuffleNet	58.9	34.6	68.1	54.1	1.71	5.1	152
<b>本文方法</b>	<b>63.2</b>	<b>39.3</b>	<b>74.2</b>	<b>57.2</b>	4.14	11.5	128

FPS 整体差异较小,这主要得益于 GPU 的并行计算能力。具体来看,YOLOv8n 由于结构简单,具有最高的推理速度;轻量化模型中,ShuffleNet 凭借高效的通道重排与并行计算特性,其 FPS 接近仅次于基线模型,而 GhostNet 由于额外的特征生成与内存访问开销,加速效果相对有限。本文方法在引入注意力机制与特征增强模块后,推理速度有所下降,但仍保持在 128 frame/s,能够满足实时检测需求,体现了较好的精度与效率平衡。

从工程应用角度看,128 frame/s 已高于常规视频流的实时处理需求,具备一定的计算冗余,可支持实际部署中的连续监测任务。综上,所提方法在水面漂浮物检测中兼顾了精度优势与推

理效率,具有良好的实际应用价值。

进一步,在中国科学院 IWHR\_Floater\_V1 公开数据集上,对现有代表性方法进行了对比实验,结果如表 2 所示。所提方法在多项关键指标上表现优异,相较于基线模型,所提方法的 mAP@50, mAP@50:95 与精确率分别提升 1.8%,0.6% 与 2.0%;与同样基于 YOLOv8 改进的 YOLOv8-HSH 算法相比,在 mAP@50 与 mAP@50:95 两项核心指标上均更优,展现出更稳定的跨数据集检测能力。虽然 RefineDet 召回率达到最高的 89.2%,但其参数量高达 23.5 M,难以满足边缘设备的轻量化部署需求,且其他性能指标均低于本文方法。

表 2 公开数据集 IWHR\_Floater\_V1 的测试结果

Tab. 2 Test results on IWHR\_Floater\_V1 dataset

Model	mAP@50/%	mAP@50:95/%	P/%	R/%	Params/M
YOLOv8n	90.1	67.1	87.1	83.6	3.01
SSD	77.5	—	85.6	79.1	22.8
RefineDet <sup>[24]</sup>	85.6	—	84.2	89.2	23.5
YOLOv8-HSH <sup>[25]</sup>	91.6	67.3	—	—	—
<b>本文方法</b>	<b>91.9</b>	<b>67.7</b>	<b>89.1</b>	<b>82.3</b>	<b>4.14</b>

表 3 展示了在公开数据集 Flow-Img 上的测试结果。相较于基线模型,本方法在 mAP@50 与 mAP@50:95 上分别提升 2.5% 与 1.2%,整体性能全面提升。与 SSD 相比,本方法以更小的模型规模获得了更优的检测精度;相较于 Rt-DETR-R50,本方法的参数量降低超过 95%,而 mAP@50 和 mAP@50:95 分别提升了 14.8% 和 6.4%。在与 YOLO 系列改进算法的对比中,本方法亦展现出良好的综合性能。与 PAR-YOLO 相比,本方法在参数量减少 44.2% 的情况下,仅 mAP@50:95 略低 1.1%;与 USD-YOLO 相比,虽然 mAP@50 降低 0.8%,但在更严格的 mAP@50:95 指标上领先 1.1%,且模型规模仅为后者三分之一,在精度与效率之间取得了良好平衡。相较于 YOLO-Float, YOLOv7-CA-Dynamic 及 YOLOv7-E6E 等模型,本方法的各项指标均有显著提升。以上结果表明,本文所提方法在 Flow-Img 数据集上有效兼顾了检测精度与模型轻量化,展现出优越的综合性能与良好的实际部署潜力。

表 3 公开数据集 Flow-Img 的测试结果

Tab. 3 Test results on Flow-Img dataset

Model	mAP@50/%	mAP@50:95/%	Params/M
YOLOv8n	82.9	45	3.01
SSD	80.6	40.3	22.8
Faster R-CNN <sup>[26]</sup>	56.2	31.7	314
Rt-DETR-R50 <sup>[27]</sup>	70.6	39.8	41.3
YOLO-Float <sup>[28]</sup>	83.3	42	91.8
YOLOv7-CD <sup>[29]</sup>	81.1	38.1	51.5
YOLOv7-E6E <sup>[30]</sup>	81.8	36.9	110.3
USD-YOLO <sup>[31]</sup>	86.2	45.1	12.35
PAR-YOLO <sup>[32]</sup>	—	47.3	7.42
<b>本文方法</b>	<b>85.4</b>	<b>46.2</b>	<b>4.14</b>

### 3.3 消融实验

#### 3.3.1 网络结构

本文基于自建数据集,在基线模型上开展系统的消融实验,以验证各项改进策略的有效性。表 4 给出了不同策略组合下的性能对比结果。整

体来看,各项改进均对模型性能产生了积极影响,验证了所提出方法的合理性与有效性。

引入 MAC-ELAN 模块后,模型在检测精度与计算效率方面均得到显著优化。其中,mAP@50 与 mAP@50:95 分别提升 2.6% 和 2.0%,同时参数量与 FLOPs 分别降低 14.6% 和 11.1%。该结果表明,MAC-ELAN 在增强特征表达能力的同时有效减少了冗余计算,实现了精度与效率的协同提升;进一步,引入 P2/P6 检测头显著增强了模型对极端尺度目标的适应能力,使 mAP@50 与精确率分别提升 1.8% 和 1.2%。然而,实验结果显示 mAP@50:95 与召回率出现轻微下降。其原因主要包括以下几个方面:首先,P2 检测头引入的高分辨率特征在强化细节表达的同时,也放大了水面倒影等背景噪声,从而对目标判别产生干扰;其次,多检测头结构增加了正样本分配的复杂性,可能导致目标在不同尺度特征层之间分配不均,影响检测稳定性;此外,不同尺度特征在语义信息与空间细节上的差异也增加了特征融合难度。尽管存在上述影响,引入 P2/

P6 检测头对于提升模型多尺度检测能力仍具有必要性,能够有效扩展模型对复杂场景中目标尺度变化的适应范围;在此基础上,C2f-FA 模块通过引入部分卷积与 EMA 注意力机制,在维持各项检测指标稳中有升的同时,有效降低了模型参数量与计算开销。其中,EMA 注意力利用自适应平均池化实现跨尺度特征建模,在一定程度上缓解了高分辨率特征带来的噪声干扰问题,从而提升了模型在复杂水面环境中的稳定性;此外,引入 WIoUv3 损失函数进一步优化了目标定位性能,使 mAP@50 与 mAP@50:95 分别提升 0.7% 和 0.2%,表明该损失函数在提高边界框回归精度方面具有积极作用。

当上述改进策略联合使用时,模型的整体性能达到最优。与基线模型相比,mAP@50 提升 4.2%,mAP@50:95 提升 3.9%,精确率与召回率分别提升 5.1% 和 2.1%。其中,mAP@50:95 的显著提升表明模型在不同 IoU 阈值下均具备更为稳定的检测能力,进一步验证了所提方法的综合优势。

表 4 不同改进点的消融实验结果

Tab. 4 Ablation analysis of different improvements

MAC-ELAN	P2/P6	C2f-FA	WIoUv3	mAP@50/%	mAP@50:95/%	P/%	R/%	Params/M	FLOPs/G
×	×	×	×	59.0	35.4	69.1	55.1	3.01	8.1
√	×	×	×	61.6	37.4	70.9	54.2	2.57	7.2
×	√	×	×	60.8	35.2	70.3	54.8	4.61	12.3
×	×	√	×	60.6	37.8	70.7	55.2	2.94	7.9
×	×	×	√	59.7	35.6	68.2	55.1	3.01	8.1
√	√	×	×	62.1	37.4	71.9	56.2	4.62	12.1
√	√	√	×	62.6	38.2	73.1	56.7	4.14	11.3
√	√	√	√	63.2	39.3	74.2	57.2	4.14	11.5

### 3.3.2 C2f-FA 特征融合模块

为验证 C2f-FA 模块中注意力机制及其关键参数设置的有效性,本文开展了包含不同通道压缩分组因子 factor 及几种常见注意力类型的消融实验,结果如表 5 所示。

首先,在 EMA 注意力的参数消融实验中,对 factor 取 4,8,16 进行对比。结果显示,Factor=4 时,mAP@50 和 mAP@50:95 分别为 62.1% 和 37.7%,Factor=16 时下降至 61.2% 和 36.7%,Factor=8 表现最佳。说明过小组虽增强通道

表 5 C2f-FA 模块的消融实验结果

Tab. 5 Ablation results of proposed C2f-FA module(%)

Model	mAP@50	mAP@50:95	P	R
Factor=4,EMA	62.1	37.7	73.1	55.3
Factor=16,EMA	61.2	36.7	71.4	55.7
Factor=8,SE	62.2	37.4	73.1	56.6
Factor=8,CA	61.9	37.2	70.6	56.9
Factor=8,CBAM	61.3	36.4	71.9	54.2
Factor=8,EMA	63.2	39.3	74.2	57.2

交互但引入冗余计算、收益有限;过大分组则削弱特征建模能力,导致精度下降,Factor=8在表达能力与计算开销间取得较优平衡。其次,在统一Factor=8条件下,将EMA替换为SE、CA和CBAM进行对比。结果表明,SE精确率接近但mAP@50:95略低;CA召回率较高但精确率下降,存在一定误检;CBAM整体性能相对较弱。相比之下,EMA在各项指标上均表现最优,该注意力机制可有效抑制复杂水面背景干扰并强化目标响应。因此,该组合为最优配置,在提升检测精度的同时兼顾模型稳定性与泛化能力。

### 3.3.3 损失函数

为验证损失函数选择的合理性,本文对WIoUv3与CIoU,SIoU等方法进行了对比实验,结果如表6所示。结果表明,WIoUv3在mAP@50,mAP@50:95及精确率等指标上均取得最优表现。相比之下,SIoU在召回率上略占优势,这主要源于其在损失构建中引入角度项与距离方向约束,将边界框回归过程分解为方向对齐与尺度逼近两个阶段,从而在优化过程中对偏移较大或质量较低的预测框施加更平滑、均衡的梯度约束,降低了对困难样本的过度抑制,提升了候选框的保留率。总体而言,WIoUv3在精度与收敛

表6 引入不同损失函数的实验结果

Tab. 6 Results of introducing different loss functions(%)

损失函数	mAP@50	mAP@50:95	P	R
CIoU	62.5	38.2	73.1	56.7
GIoU	61.8	37.6	70.5	56.0
DIoU	61.2	36.4	68.2	57.0
SIoU	61.1	36.3	68.1	57.5
EIoU	61.5	37.7	72.5	56.2
WIoUv1	62.4	37.5	69.1	57.2
WIoUv2	61.8	37.4	70.3	56.8
WIoUv3	63.2	39.3	74.2	57.2

稳定性方面更具优势,但在召回率上存在一定权衡,反映了不同损失函数在优化目标与性能侧重点上的差异。考虑到水面漂浮物检测场景中存在倒影干扰与小目标等问题,对定位精度与误检抑制能力要求更高,WIoUv3在综合性能上更适用于本文模型,同时也有助于提升复杂背景下的检测稳定性与泛化能力。

### 3.4 可视化对比结果

图6和图7分别展示了小目标密集场景以及复杂背景与多尺度目标场景下的检测结果对比。

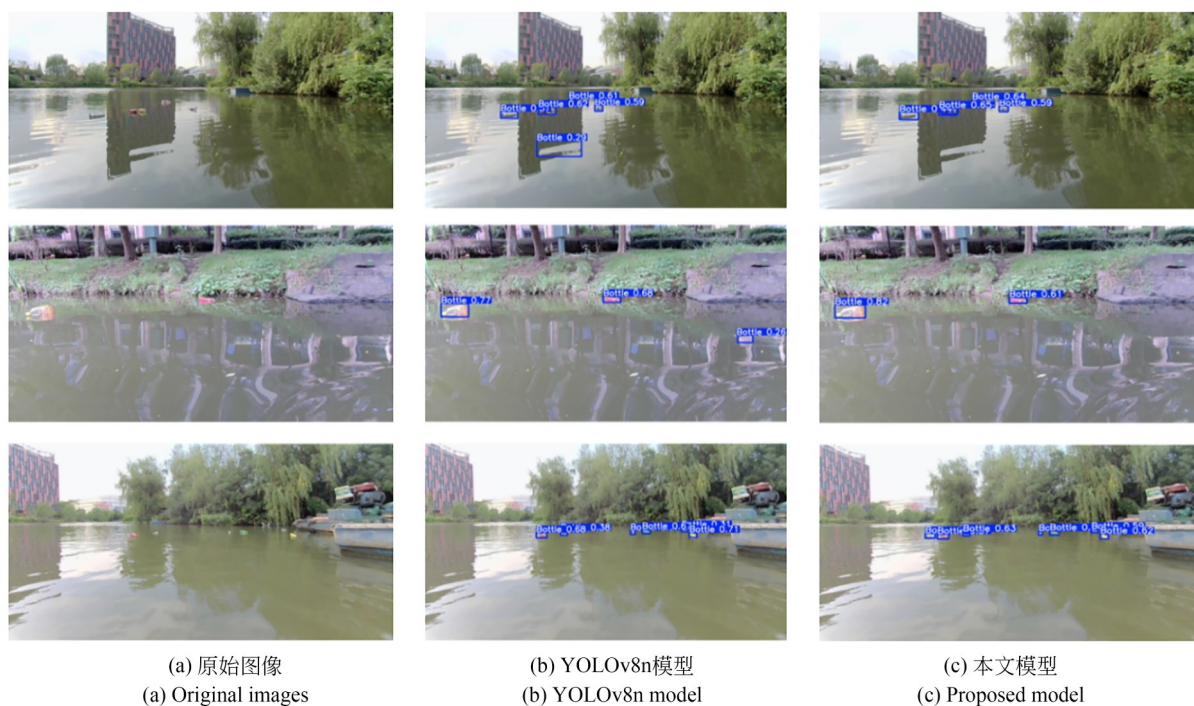


图6 小目标较多的场景下不同算法的检测结果

Fig. 6 Detection results of different algorithms in scenes with many small objects

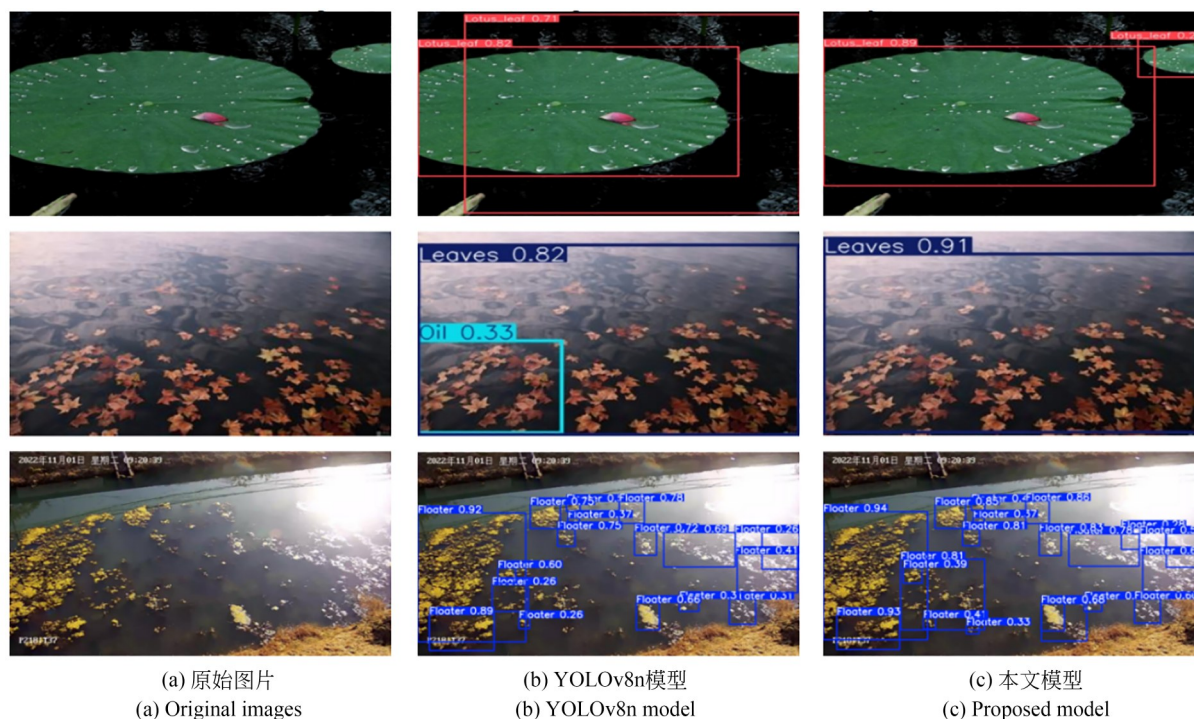


图 7 复杂背景或目标尺度大的场景下不同算法的检测结果

Fig. 7 Detection results of different algorithms in scenes with complex backgrounds or large target scales

总体来看,相较于基准模型 YOLOv8n,本文方法在多种复杂水面环境中均表现出更优的性能。

在小目标场景中, YOLOv8n 存在一定漏检, 并易受水面倒影干扰产生误检, 而本文方法能够更完整、准确地识别目标, 且预测置信度更高, 体现出更强的小目标表征能力与抗干扰能力。在复杂背景及多尺度场景中, YOLOv8n 对尺度变化较大的目标识别能力有限, 而本文方法能够稳定检测不同尺度目标, 并有效抑制背景干扰, 检测结果更加可靠。综合来看, 所提模型在小目标检测、多尺度适应及复杂背景抑制方面均具有明显优势, 表现出更好的泛化能力与稳定性。

## 4 结 论

针对水面漂浮物检测中背景复杂、目标尺度差异大等挑战, 本文在 YOLOv8n 基础上提出一套轻量化改进方案。首先, 设计 MAC-ELAN 模块嵌入主干与颈部网络, 增强多尺度特征表达能

力并缓解深层梯度丢失, 有效提升对模糊边界的判别能力; 其次, 引入 P2 与 P6 检测头构建五层检测架构, 强化对极小目标与宏观污染区域的同步感知; 进一步, 提出 C2f-FA 模块以部分卷积与 EMA 注意力融合实现高效特征提取, 在控制参数数量的同时保持模型精度; 最后, 采用 WIoUv3 损失函数增强定位稳定性, 提升复杂背景下的回归性能。实验结果表明, 本文方法在自建与公共数据集上均表现优异, 验证了各改进策略的有效性与协同作用。该方法为水面漂浮物检测提供了可靠的技术路径, 也为复杂环境下的目标检测任务提供了可扩展的改进框架。后续工作将聚焦于模型轻量化与极端天气条件下的检测性能提升。

### 作者贡献说明:

杨天麒: 方法论, 初稿撰写, 实验设计, 数据整理和分析;

王 勇: 论文构思与写作指导, 提供资源。

### 参考文献:

[1] JIA T L, KAPELAN Z, DE VRIES R, *et al.*

Deep learning for detecting macroplastic litter in water bodies: a review [J]. *Water Research*, 2023,

- 231; 119632.
- [2] ZUO J J, WU Y F. Recognition method for abnormal water surface based on SOM[J]. *Proceedings of the 2nd International Conference on Modelling, Identification and Control*, 2015.
- [3] SUN X, DENG H, LIU G H, *et al.* Combination of spatial and frequency domains for floating object detection on complex water surfaces [J]. *Applied Sciences*, 2019, 9(23): 5220.
- [4] LI N, HUANG H, WANG X Y, *et al.* Detection of floating garbage on water surface based on PC-net [J]. *Sustainability*, 2022, 14(18): 11729.
- [5] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single Shot MultiBox Detector[M]. *Computer Vision-ECCV 2016*. Cham: Springer International Publishing, 2016: 21-37.
- [6] JIANG P Y, ERGU D, LIU F Y, *et al.* A review of yolo algorithm developments[J]. *Procedia Computer Science*, 2022, 199: 1066-1073.
- [7] CHEN R F, WU J, PENG Y, *et al.* Detection and tracking of floating objects based on spatial-temporal information fusion[J]. *Expert Systems with Applications*, 2023, 225: 120185.
- [8] LIN F, HOU T, JIN Q N, *et al.* Improved YOLO based detection algorithm for floating debris in waterway[J]. *Entropy*, 2021, 23(9): 1111.
- [9] WEN G, LI S B, LIU F C, *et al.* YOLOv5s-CA: a modified YOLOv5s network with coordinate attention for underwater target detection [J]. *Sensors*, 2023, 23(7): 3367.
- [10] YANG X S, ZHAO J Y, ZHAO L, *et al.* Detection of river floating garbage based on improved YOLOv5[J]. *Mathematics*, 2022, 10(22): 4366.
- [11] ZHANG X H, MIN C Z, LUO J W, *et al.* YOLOv5-FF: detecting floating objects on the surface of fresh water environments[J]. *Applied Sciences*, 2023, 13(13): 7367.
- [12] LI H, YANG S P, ZHANG R, *et al.* Detection of floating objects on water surface using YOLOv5s in an edge computing environment[J]. *Water*, 2023, 16(1): 86.
- [13] WANG H R, CHENG H, ZHANG J T. Faster-PGYOLO: an efficient framework for floating debris detection in inland waters [J]. *The Visual Computer*, 2025, 41(7): 5087-5104.
- [14] HUANGFU Z M, LI S Q, YAN L H. Ghost-YOLO v8: an attention-guided enhanced small target detection algorithm for floating litter on water surfaces [J]. *Computers, Materials & Continua*, 2024, 80(3): 3713-3731.
- [15] 孙世政, 何玲玲, 郑帅, 等. 改进 YOLOv8n 的复杂环境下垃圾轻量化检测[J]. *光学精密工程*, 2025, 33(12): 1984-1998.
- SUN SH ZH, HE L L, ZHENG SH, *et al.* Improved lightweight garbage detection method for YOLOv8n in complex environments [J]. *Optics and Precision Engineering*, 2025, 33(12): 1984-1998. (in Chinese)
- [16] 张彪, 朱振洋, 许家忠. 面向自主水下机器人的深海污染物检测[J]. *光学精密工程*, 2025, 33(11): 1803-1817.
- ZHANG B, ZHU ZH Y, XU J ZH. Deep-sea pollutant detection for autonomous underwater robots [J]. *Optics and Precision Engineering*, 2025, 33(11): 1803-1817. (in Chinese)
- [17] TONG Z J, CHEN Y H, XU Z W, *et al.* Wise-IoU: bounding box regression loss with dynamic focusing mechanism [EB/OL]. 2023: *arXiv*: 2301.10051. <https://arxiv.org/abs/2301.10051>.
- [18] DING X H, ZHANG X Y, MA N N, *et al.* RepVGG: making VGG-style ConvNets great again[C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 13728-13737.
- [19] CHEN J R, KAO S H, HE H, *et al.* Run, don't walk: chasing higher FLOPS for faster neural networks[C]. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 12021-12031.
- [20] CHENG Y W, ZHU J N, JIANG M X, *et al.* FloW: a dataset and benchmark for floating waste detection in inland waters [C]. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 10933-10942.
- [21] QIAO G C, YANG M X, WANG H. An annotated dataset and benchmark for detecting floating debris in inland waters [J]. *Scientific Data*, 2025, 12: 385.
- [22] HAN K, WANG Y H, TIAN Q, *et al.* GhostNet: more features from cheap operations [EB/OL]. 2019: *arXiv*: 1911.11907. <https://arxiv.org/abs/1911.11907>.

- org/abs/1911.11907.
- [23] ZHANG X Y, ZHOU X Y, LIN M X, *et al.* ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 6848-6856.
- [24] ZHANG S F, WEN L Y, BIAN X, *et al.* Single-shot refinement neural network for object detection [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 4203-4212.
- [25] SONG L L, DENG H X, HAN J F, *et al.* Improved aerial surface floating object detection and classification recognition algorithm based on YOLOv8n[J]. *Sensors*, 2025, 25(6): 1938.
- [26] GIRSHICK R. Fast R-CNN[C]. 2015 *IEEE International Conference on Computer Vision (ICCV)*. December 7-13, 2015. Santiago, Chile. IEEE, 2015: 1440-1448.
- [27] ZHAO Y A, LV W Y, XU S L, *et al.* DETRs beat YOLOs on real-time object detection [C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 16965-16974.
- [28] LI K, WANG Y N, HU Z M. Improved YOLOv7 for small object detection algorithm based on attention and dynamic convolution [J]. *Applied Sciences*, 2023, 13(16): 9316.
- [29] LI Y, WANG R C, GAO D X, *et al.* A floating-waste-detection method for unmanned surface vehicle based on feature fusion and enhancement [J]. *Journal of Marine Science and Engineering*, 2023, 11(12): 2234.
- [30] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 7464-7475.
- [31] DENG H Q, ZHANG S Z, WANG X D, *et al.* USD-YOLO: an enhanced YOLO algorithm for small object detection in unmanned systems perception[J]. *Applied Sciences*, 2025, 15(7): 3795.
- [32] LI N, WANG M L, HUANG H, *et al.* PAR-YOLO: a precise and real-time YOLO water surface garbage detection model[J]. *Earth Science Informatics*, 2025, 18(1): 135.

## 作者简介:



杨天麒(2001—),男,硕士研究生,主要研究方向为机器视觉与深度学习应用。E-mail: 993550572@qq.com

## 通讯作者:



王 勇(1980—),男,湖北仙桃人,教授,2009年于华中科技大学获得博士学位,主要研究方向为物联网监测、人工智能应用。E-mail: wy112708@163.com