

DOI:10.19651/j.cnki.emt.2519994

改进 YOLOv11n 的水面垃圾检测算法^{*}王海群¹ 谢伟民¹ 晁 帅² 于海峰¹

(1. 华北理工大学电气工程学院 唐山 063200; 2. 华北理工大学招生就业处 唐山 063200)

摘要: 为应对水波干扰、光线反射及漂浮物遮挡等复杂水面背景下目标检测精度下降的问题,提出一种改进 YOLOv11n 的水面垃圾检测算法。首先,利用 FCM 模块对 C3k2 中的 Bottleneck 进行重新设计,构建 C3k2-FCM 模块,缓解下采样过程中物体空间位置信息丢失的问题;其次,引入具有不同膨胀率的共享卷积层并结合空间与通道协同注意力机制 SCSA,构建 FPSC-SCSA 模块取代 SPPF,增加对关键区域的关注能力,减少关键信息的丢失;再次,将颈部聚合网络替换为多尺度特征融合网络并嵌入 U-Net V2 的 SDI 模块构建 BIFPN-V2S 模块,扩展模型感受野并强化全局上下文信息交互能力;最后,构建 SGSV-IoU 损失函数,提高模型对边界形状和细节的刻画精度,优化对不规则漂浮物的定位效果。实验结果表明:改进后的 YOLOv11n 算法在自建水面垃圾数据集上相较原算法 mAP50 提高了 2.8%,参数量和计算量分别减少了 0.81 M 和 0.1 GFLOPs,证明了改进算法的有效性。

关键词: 水面垃圾;YOLOv11n;损失函数;SCSA 注意力机制;共享卷积层;多尺度特征融合

中图分类号: TN911.73;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.40

Improved YOLOv11n-based water surface garbage detection algorithm

Wang Haiqun¹ Xie Weimin¹ Chao Shuai² Yu Haifeng¹

(1. School of Electrical Engineering, North China University of Science and Technology, Tangshan 063200, China;

2. Admissions and Employment Office, North China University of Science and Technology, Tangshan 063200, China)

Abstract: To address the decline in detection accuracy caused by complex water surface backgrounds such as wave interference, light reflection, and object occlusion, an improved YOLOv11n-based water surface garbage detection algorithm is proposed. First, the Bottleneck structure in C3k2 is replaced with the FCM module to construct the C3k2-FCM module, mitigating the loss of spatial position information during downsampling. Second, a FPSC-SCSA module is designed by introducing shared convolutional layers with different dilation rates and the SCSA mechanism to replace the SPPF module, enhancing the model's focus on key regions and reducing the loss of crucial information. Third, a BIFPN-V2S module is developed by replacing the neck aggregation network with a multi-scale feature fusion network embedded with the SDI module from U-Net V2, expanding the receptive field and strengthening global contextual interactions. Finally, an SGSV-IoU loss function is designed to improve boundary shape and detail representation, enhancing the localization of irregular floating objects. Experimental results on a self-built water surface garbage dataset show that, compared with the original YOLOv11n model, the improved algorithm increases mAP50 by 2.8%, while reducing parameters and computation by 0.81 M and 0.1 GFLOPs, respectively, demonstrating the effectiveness of the proposed method.

Keywords: water surface garbage; YOLOv11n; loss function; SCSA attention mechanism; shared convolutional layer; multi-scale feature fusion

0 引言

近年来,塑料垃圾对淡水生态系统造成的污染问题日益严重^[1],已成为全球环境治理的焦点之一,中国同样深受

其害,面临严峻的生态压力与治理挑战。为了有效监控和管理淡水垃圾的问题,开发高效且精准的水面垃圾检测技术变的至关重要。

传统的巡检方法通常依赖人工打捞或其他辅助设备收

收稿日期:2025-10-10

^{*} 基金项目:河北省自然科学基金(D2024209006)、河北省教育厅科学研究项目(QN2024147)资助

集水面垃圾,却因难以掌握垃圾分布位置,存在作业效率低、漏捞严重、人力与时间成本高等问题。随着深度学习技术的迅猛发展,基于目标检测算法的水面垃圾检测方法逐渐成为水面垃圾检测的重要模式,不仅提高了巡检效率,还极大减少了人力成本。

目前基于深度学习的目标检测算法主要分为两阶段算法(如 Faster R-CNN^[2]、Mask R-CNN^[3]等)和单阶段算法(YOLO^[4]系列、TOOD^[5]等)。在两阶段检测算法中,Bai 等^[6]对 Faster R-CNN 进行了改进,引入 ResNet50 作为特征提取骨干网络,并优化锚框生成与训练策略,提升了对海洋浮动垃圾的检测精度。Ma 等^[7]基于 Mask R-CNN 算法,在结合强监督与弱监督策略下又引入对比语言-图像预训练(contrastive language-image pretraining, CLIP)模型进行跨模态特征对齐,提高对目标检测的精度。但此类算法因模型参数众多且计算开销较大,限制了其在资源受限的边缘设备上的部署与应用。相较之下,以 YOLO 为代表的单阶段检测算法因其轻量化的模型结构和较低的计算复杂度,使其能够更好的满足实际需求。Li 等^[8]针对水面漂浮垃圾检测,改进了 YOLOv3 算法结构,将原有的三尺检测机制调整为两尺度检测,同时对锚框参数进行创新设计与优化,提高模型对目标的定位与识别检测精度。Doherty 等^[9]基于 YOLOv5 模型框架,改进特征融合部分,使用双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)替换原有的 PANet 结构,显著提高了多尺度特征融合的效率。石露露等^[10]针对明渠漂浮垃圾检测,提出一种改进的 YOLOv5s 算法。首先通过多种数据增强方法扩充数据集,以提高模型在不同场景下的泛化能力,随后在 YOLOv5s 结构中引入加权双向特征融合网络,优化多尺度特征融合过程,提升对小目标及复杂背景下的检测精度。上述算法虽然在基础模型上进行了多项改进,并取得了一定的性能提升,但在实际水面垃圾检测中,仍易受到水波干扰、光线反射及漂浮物遮挡等复杂水面环境影响,导致部分目标出现漏检与误检现象。

针对上述问题,本文提出一种改进 YOLOv11n 的水面垃圾检测算法。首先,通过在主干网络里引入改进的 C3k2 模块^[11],增强主干网络特征提取与表达能力;其次,将空间与通道协同注意力机制(shared multi-semantic spatial attention, SCSA)^[12]引入多膨胀率共享卷积层,用以重构快速空间金字塔池化(spatial pyramid pooling fast, SPPF)模块,减少局部细节信息丢失;随后,引入改进后的双向特征融合网络(bidirectional feature pyramid network, BiFPN)^[13]取代 YOLOv11 原始的路径聚合网络(path aggregation network, PANet)^[14]作为 Neck 部分的特征融合方式,实现多尺度特征之间的充分交互与补偿;最后,采用融合广义表面损失(generalized surface loss, GSL)^[15]和方差感知焦点损失(variance-aware focal loss, varifocal loss)^[16]的 Shape-IoU(shape intersection over union)^[17]取代原损失函数,提高模型

边界回归能力,同时加速网络收敛速度。

1 YOLOv11 的网络结构

YOLOv11 是 Ultralytics 推出的最新一代目标检测框架,其延续 YOLO 系列的模块化设计将模型划分为 n、s、m、l 和 x 5 种规格。YOLOv11n 凭借轻量化的网络结构和低计算成本,在实际应用中表现出较高应用价值。整体网络构架可划分为 3 大模块:主干网络(Backbone)、颈部网络(Neck)以及输出网络(Head),网络结构如图 1 所示。

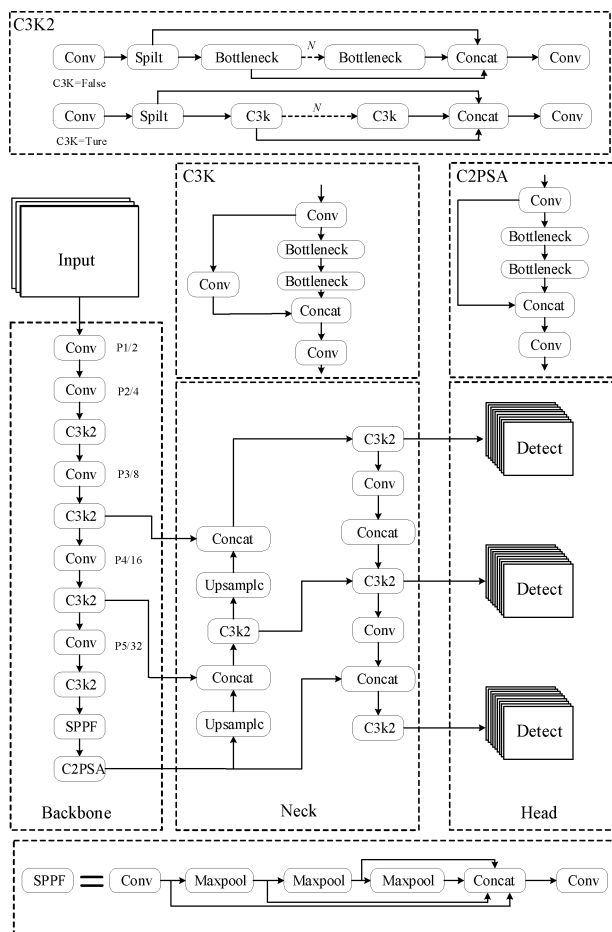


图 1 YOLOv11n 网络结构

Fig. 1 YOLOv11n network structure

YOLOv11n 在主干网络(Backbone)中引入 C3k2 模块,该模块在继承 C2f 特征融合思想的同时,增加灵活切换机制。当参数 C3K 设置为 True 时,传统的瓶颈结构被更高效的 C3 模块替代,从而提高特征提取能力与网络适应性。此外,YOLOv11 还提出了 C2PSA 模块,该模块结合了并行多头注意力(parallel spatial attention, PSA)机制与前馈神经网络,使其能在复杂背景下突出目标特征,抑制无关区域的干扰。Neck 部分相较于 YOLOv8 变化较小,仍沿用原始的多尺度特征融合机制,但却将原有的标准卷积替换为通道数为 1/4 的轻量化卷积,以降低计算开销。Head 部分采用完全解耦结构,将分类与回归任务分开处

理,避免两者在联合优化过程中产生冲突,提高模型的收敛速度与检测精度。

2 改进 YOLOv11n 算法

2.1 构建 FCM-C3k2 模块

在 YOLOv11n 中,C3k2 模块中 Bottleneck 结构先通过的 1×1 卷积降低通道数,再由 3×3 卷积进行特征提取,最后恢复通道。其利用残差连接,不仅缓解了梯度退化和

信息丢失的问题,还有效降低了计算量。然而,通过压缩通道与扩张的方式在实现特征降维与复原的过程中,会导致部分细节信息丢失,降低模型在小目标检测中对局部特征的敏感性。为突破 Bottleneck 结构的局限性,本文在 C3K2 模块中引入特征互补映射模块(feature complementary mapping module,FCM)^[18]替代 Bottleneck,以在保持高效特征提取能力前提下,进一步增强模型对局部细节特征的捕捉能力。FCM 模块结构如图 2 所示。

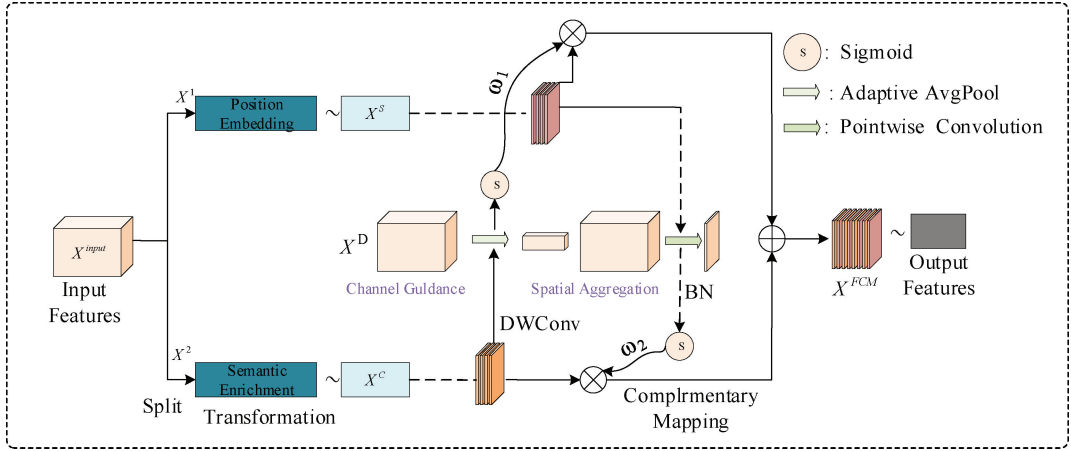


图 2 FCM 模块工作原理

Fig. 2 Schematic of the operating principle of the FCM module

FCM 通过多维度的特征分解与交互机制,实现浅层空间信息与深层语义信息的高效融合,本研究将输入特征通道按特征分割比例 α 进行分割,得到 αC 与 $(1-\alpha)C$,为增强小目标的检测性能,取 $\alpha = 0.25$,以在保证语义特征表达能力的同时最大限度保留空间细节信息。再将分割后得到的两部分特征 X^1, X^2 分别送入不同的分支进行处理,得到特征信息 X^C 与浅层空间位置信息 X^S ,该转换过程由式(1)表示:

$$(X^C, X^S) = \phi_1(X^1, X^2) \quad (1)$$

式中: ϕ_1 表示学习空间信息与语义信息间的映射关系。

但当前获得的特征 X^C 和 X^S 有效却呈离散状态,这可能导致目标特征匹配不精确,因此要对其进行互补映射,通过通道交互和空间交互,分别将重要信息映射到另一条分支,实现特征的互补融合,得到的通道信息权重 ω_1 和空间信息权重 ω_2 。再将 ω_1, ω_2 分别映射到包含 X^S 和 X^C 的特征上,然后将两个分支连接起来,得到包含空间和语义关系双重映射特征的 X^{FCM} ,其计算公式为:

$$\omega_1 = \mathbb{R} \left(\frac{1}{H \times W} \sum_{i=0}^H \sum_{j=0}^W X^D(i, j) \right) \quad (2)$$

$$\omega_2 = \mathbb{R}(F(X^S)) \quad (3)$$

$$X^{FCM} = (X^S \otimes \omega_2) \oplus (X^C \otimes \omega_1) \quad (4)$$

式中: X^D 表示深度卷积后的输出结果, H 和 W 分别表示特征图的高度和宽度, F 表示带空间聚合的卷积映射, $\mathbb{R}$ 代表激活函数, $\oplus$ 表示逐元素加法, $\otimes$ 表示逐元素乘法。

通过在特征提取模块上引入 FCM,弥补了 Bottleneck 结构难以捕捉局部细节的局限性。因此 C3k2-FCM 模块能够在原有 C3k2 的基础上,进一步增强了网络对多尺度细节特征的表达能力。实现了对目标边缘与纹理更精确的刻画,为后续特征融合与检测头预测提供了更具判别力的特征支撑。

2.2 聚合金字塔共享模块 FPSC-SCSA

SPPF 作为主干网络的重要组成部分,其利用空间金字塔池化捕捉不同尺度的特征,并进行多尺度信息融合。但是,由于水面垃圾目标形状多变,类别复杂,传统 SPPF 模块固定的池化策略在特征提取方面存在局限性。针对这一问题,本文对 SPPF 进行优化改进,引入多膨胀率的特征金字塔共享卷积(FPSC),并结合 SCSA,构建出多尺度特征金字塔共享卷积模块 FPSC-SCSA。

在 FPSC-SCSA 模块中,采用多膨胀率卷积结构以实现不同尺度水面目标的精细化建模。卷积核尺寸固定为 3×3 ,膨胀率设置为 $d = \{1, 3, 5\}$ 。其中, $d = 1$ 的分支侧重捕获小尺度目标的细粒度纹理与边缘信息; $d = 3$ 的分支关注中等尺度目标的局部结构特征; $d = 5$ 的分支则通过扩大感受野提取大尺度目标的全局语义信息。多膨胀率卷积的并行设计使模型能够在共享卷积权重的条件下高效融合不同感受野的特征表征,在空间细节与语义上下文之间建立动态平衡,有效缓解了传统 SPPF 模块在复杂水面场景下对多尺度目标响应不足的问题。

空间与通道协同注意力机制 SCSA 能够利用不同特

征通道之间固有的空间语义差异,指导关键特征学习。同时,通过缓解多语义特征间的语义不一致性并结合信息融

合策略,实现不同语义层次和通道特征的优化整合,提高对关键特征的提取能力。其结构如图 3 所示。

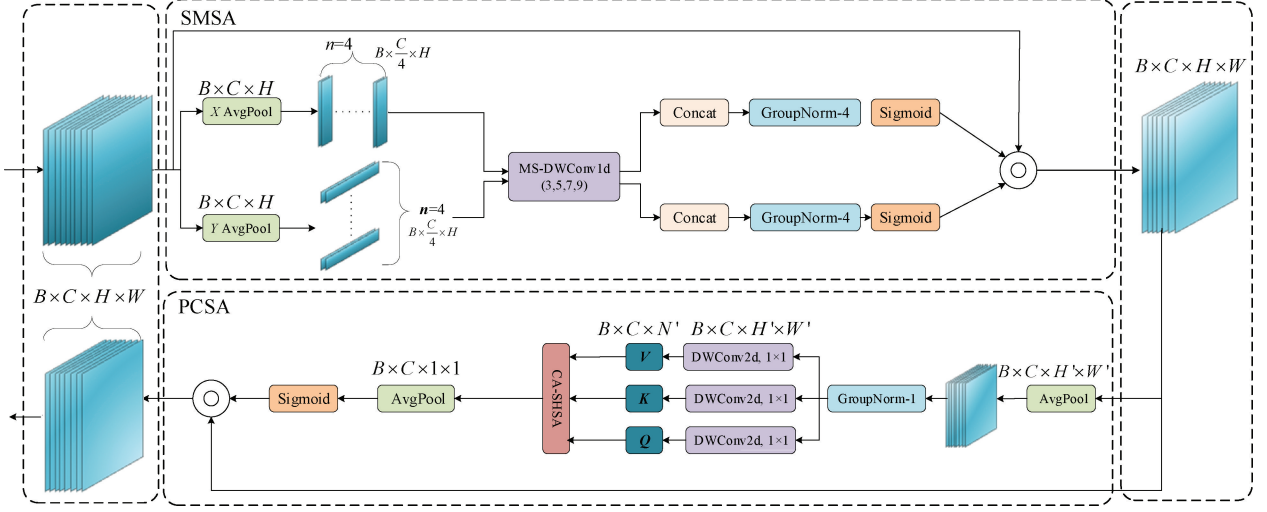


图 3 SCSA 注意力机制原理

Fig. 3 Schematic of the SCSA attention mechanism

图 3 所示结构中,SCSA 将输入信息先传入共享多语义空间注意力(shared multi-semantic spatial, SMSA)机制,其原理是将输入特征沿高、宽两种维度进行分解,并进一步划分为 \$K\$ 个独立子特征以表征不同语义的空间信息。随后将子特征拼接并通过组归一化(group normalization, GN)进行独立标准化,再通过 Sigmoid 激活函数生成空间注意力图,对特定空间区域进行选择性的增强或抑制,获得语义感知优化的输出特征。SMSA 的定义式为:

$$Attn_H = \sigma(GH_H^K \text{Concat}(\tilde{X}_H^1, \tilde{X}_H^2, \dots, \tilde{X}_H^K)) \quad (5)$$

$$Attn_W = \sigma(GH_W^K \text{Concat}(\tilde{X}_W^1, \tilde{X}_W^2, \dots, \tilde{X}_W^K)) \quad (6)$$

$$\text{SMSA}(X) = X_S = Attn_H \times Attn_W \times X \quad (7)$$

式中: \$\sigma\$ 表示 Sigmoid 归一化函数, \$GH_W^K(\cdot)\$ 和 \$GH_H^K(\cdot)\$ 分别表示在宽度 \$W\$ 和高度 \$H\$ 上划分为 \$K\$ 个组的组归一化。

再将优化后的输出特征输入至渐进通道自注意力(progressive channel-wise self-attention, PCSA)机制,该机制采用输入感知的自注意机制进一步优化通道特征,能有效缓解不同语义层次的差异。PCSA 的定义式为:

$$X_P = \text{Pool}_{(7,7)}^{(H,W) \rightarrow (H',W')}(X_S) \quad (8)$$

$$X_{Attn} = \text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{C}}\right)\mathbf{V} \quad (9)$$

$$\text{PCSA}(X_S) = X_C = X_S \times \sigma(\text{Pool}_{(H',W') \rightarrow (1,1)}^{(H,W) \rightarrow (H',W')}(X_{Attn})) \quad (10)$$

式中: \$\text{Pool}_{(K,K)}^{(H,W) \rightarrow (H',W')}(\cdot)\$ 表示池化操作,其卷积核大小为 \$K \times K\$,用于将特征图的分辨率从 \$(H, W)\$ 缩放到 \$(H', W')\$。\$\mathbf{Q}\$、\$\mathbf{K}\$ 和 \$\mathbf{V}\$ 分别表示通过线性变换得到的查询特征、键特征和值特征矩阵, \$\sqrt{C}\$ 表示通道数的缩放系数。

FPSC-SCSA 模块通过引入空间与通道协同注意力机

制 SCSA,实现了对特征图中多尺度关键信息的自适应关注,使得模型能准确的关注多个关键区域,最大限度地减少了关键信息的丢失,为颈部网络提供丰富的特征信息,从而进一步提升模型的检测性能。

2.3 特征融合网络 BIFPN-V2S

YOLOv11n 颈部采用路径聚合网络 PANet 进行特征融合,该结构虽增强了不同层级之间的信息交互,但过度依赖固定的上采样与拼接方式,导致灵活性不足,且对非相邻层级特征的利用有限,易造成目标信息丢失。多尺度特征融合网络 BIFPN 通过双向连接与加权融合机制,在自顶向下和自底向上的特征流之间引入跨尺度连接,有效解决了 PANet 在非相邻层级特征利用上的局限性。BIFPN 结构如图 4(a)所示。

为了实现更高效的特征融合,本文在采用 BIFPN 的基础上,将网络原有的拼接(Concat)操作替换为 U-Net V2 中的 SDI^[19]模块。SDI 的结构原理如图 4(b)所示,该模块基于编码器生成的分层特征映射,在每一层特征 \$i\$ 上整合空间注意力与通道注意力机制,使特征能够兼顾局部空间结构与全局通道信息。在通过 \$1 \times 1\$ 卷积将 \$f_i^1\$ 通道数压缩到 \$c\$,得到 \$f_i^2 \in \mathbb{R}^{H_i \times W_i \times c}\$。接着将该特征图输入到特征恢复模块,并以 \$f_i^2\$ 作为基准,在每个第 \$j\$ 层对特征图进行缩放操作,使其与 \$f_i^2\$ 保持相同的尺寸,最后将调整后的特征图 \$f_{ij}^3\$ 应用 \$3 \times 3\$ 卷积进行平滑处理,得到特征图 \$f_{ij}^4\$。主要计算公式为:

$$f_i^1 = \phi_i^1(\phi_i^1(f_i^0)) \quad (11)$$

$$f_{ij}^3 = \begin{cases} D(f_j^2, (H_i, W_i)), & j < i \\ I(f_j^2), & j = i \\ U(f_j^2, (H_i, W_i)), & j > i \end{cases} \quad (12)$$

$$f_{ij}^4 = \theta_{ij}(f_{ij}^3) \quad (13)$$

式中: f_i^1 表示第 i 层处理过的特征映射; ϕ_i^c 和 ϕ_i^s 分别表示该层对应的空间和通道注意力权重参数, D 、 I 和 U 分别表

示自适应平均池化、恒等映射和双线性插值函数, H_i 、 W_i 表示特征图的长度和宽度; θ_{ij} 表示平滑卷积的卷积参数; f_{ij}^4 表示第 i 层的第 j 个平滑后的特征图。

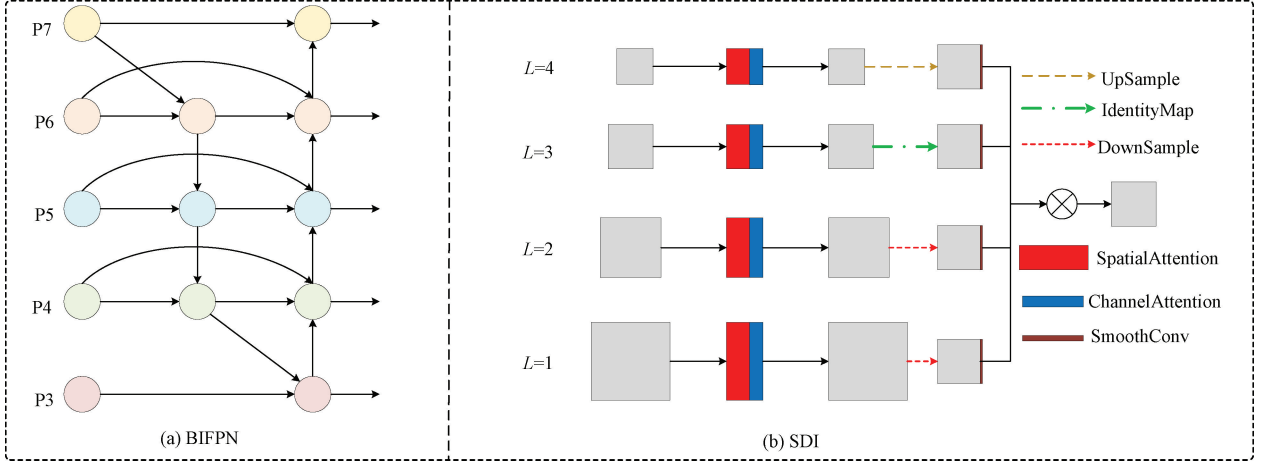


图 4 BIFPN 模块与 SDI 模块工作原理

Fig. 4 Schematic of the working principles of the BIFPN and SDI modules

为进一步提高特征融合效果,将所有特征图调整至统一尺寸,并对这些特征图施加逐元素乘积(Hadamard)操作,实现对第 i 层特征的增强,使其在不增加额外参数开销的前提下扩展感受野、增强全局上下文信息交互。同时对多尺度特征选择性融合,减少冗余干扰,保留关键信息。所以,将改进后的 BIFPN-V2S 模块引入模型,能使模型在复杂水面环境及多种干扰条件下表现出更高的鲁棒性和检测精度。

2.4 损失函数 SGSV-IoU

YOLOv11n 在边界框回归中采用 CIoU (complete intersection over union) 损失函数,该方法是在 IoU 基础上引入中心点距离和宽高比约束,使预测框在位置、重叠度和形状上更接近真实框,并结合分布式焦点损失 (distribution focal loss, DFL) 以提升对边界的预测精度^[20]。但在水面垃圾检测中,CIoU 面对小目标或长宽差异较大的目标时,目标长宽比例的梯度可能过小,导致优化不足^[21]。同时面对重叠的水面垃圾,CIoU 对目标边缘和轮廓的区分能力较弱,可能导致检测框偏离实际目标^[22]。为解决上述问题,本文引入以形状交并比为核心设计的 Shape-IoU 作为新的边界损失函数,原理如图 5 所示。

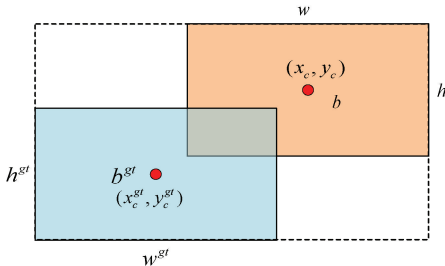


图 5 Shape-IoU 工作原理

Fig. 5 Principle of Shape-IoU

在损失计算中引入 Shape-IoU,通过在传统 IoU 指标中引入尺度因子及与真实框形状相关的权重系数,实现对水平方向和垂直方向的偏差的自适应调节。该设计在权衡预测框与真实框的重叠程度的同时,进一步融合边界框的尺寸比例与形状差异,对回归误差进行多维约束。通过将重叠度与形状信息协同建模,Shape-IoU 能有效提升边界框的定位精度。

Shape-IoU 是主要由 3 部分构成:IoU、中心距离 $distance^{shape}$ 和形状惩罚项 Ω^{shape} ,其中 IoU 用来衡量预测框和真实框的相似度,中心距离是衡量两者中心的几何距离,形状惩罚项通过对边界框形状进行建模,量化其对回归精度的影响。相关表达式为:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (14)$$

$$distance^{shape} = hh \times \left(\frac{x_c - x_c^{gt}}{c} \right)^2 + ww \times \left(\frac{y_c - y_c^{gt}}{c} \right)^2 \quad (15)$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-wt})^\theta, \theta = 4 \quad (16)$$

$$\begin{cases} w_w = hh \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ w_h = ww \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (17)$$

式中: B 表示预测框, B^{gt} 表示真实框, ww 和 hh 分别表示水平方向和垂直方向的权重参数, x_c^{gt} 和 y_c^{gt} 分别对应真实框中心坐标, c 为缩放因子, w^{gt} 和 h^{gt} 分别表示真实框的宽度和高度。基于上述参数,Shape-IoU 边界框回归损失函数表达式为:

$$L_{shape-IoU} = 1 - IoU + distance^{shape} + 0.5 \times \Omega^{shape} \quad (18)$$

鉴于 Shape-IoU 在水面垃圾检测任务中对小目标尺度敏感性不足,以及对不规则目标适应性有限,导致模型在不规则漂浮物定位上存在一定偏差。为此,本文将 GSL 和 Varifocal Loss 引入边界回归与分类优化中进一步优化 Shape-IoU。其中,GSL 通过全局与局部几何约束增强模型对边界形状和细节的刻画能力,改善对不规则漂浮物的定位效果^[23]。Varifocal Loss 利用自适应正负样本加权机

制,缓解样本不平衡带来的训练偏差,提高模型在复杂水面场景下的分类准确性。基于此,本文将改进后的损失函数命名为 SGSV-IoU,定义为:

$$L_{SGSV-IoU} = L_{Shape-IoU} (1 + \lambda_1 L_{GSL}) + \lambda_2 L_{Varifocal} \quad (19)$$

经试验表明,当 λ_1 为 0.5, λ_2 为 1 时模型在检测精度最佳。基于上述改进策略,本文最终构建了优化后的 YOLOv11n 网络结构,其整体框架如图 6 所示。

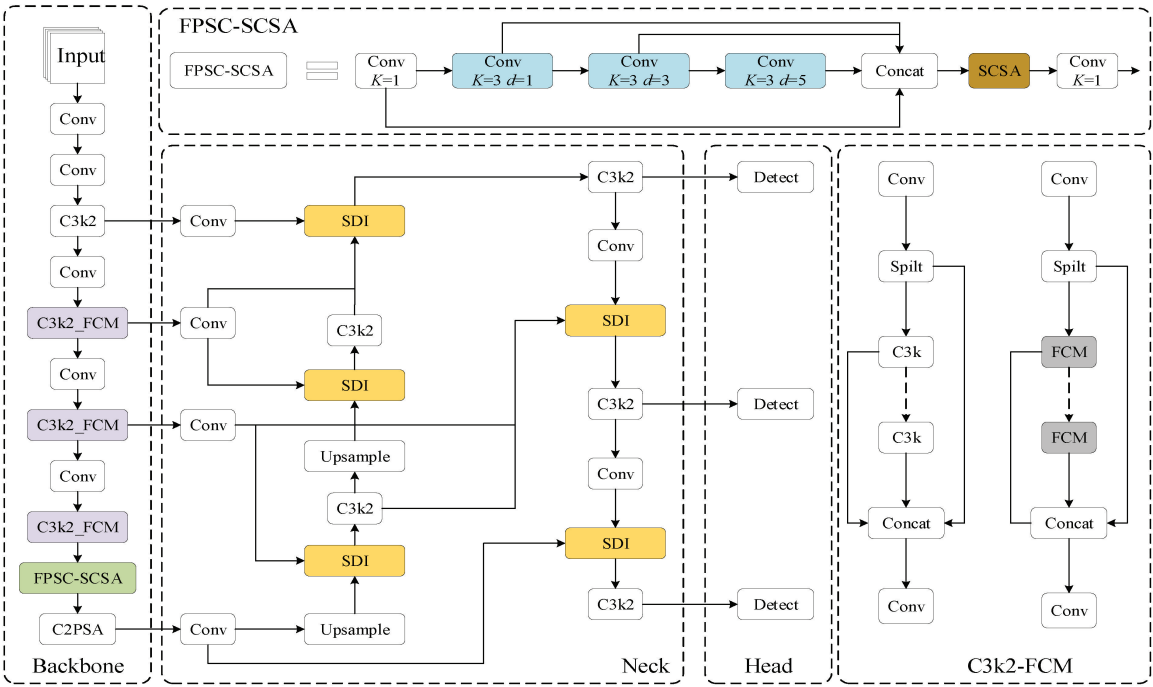


图 6 改进后的 YOLOv11n 网络结构
Fig. 6 The improved YOLOv11n network architecture

3 实验结果与分析

3.1 数据集介绍

本实验使用的水面垃圾图像数据来源于无人机巡航拍摄和人工采集,共包含 4 274 张水面垃圾图片,涵盖 8 类目标:瓶子(bottle)、杂草(grass)、树枝(branch)、牛奶盒(milk-box)、塑料袋(plastic-bag)、塑料垃圾(plastic-garbage)、皮球(ball)及树叶(leaf)。部分目标类别示例如图 7 所示。为保证数据的随机性,所有图像在训练前均进行了随机打乱,并依据 PASCAL VOL 的标注规范,使用 LabelImg 工具进行精确标注^[24]。由于皮球类别仅有 28 张样本,远低于其他类别,故本实验对其进行数据增强处理,以平衡类别分布。数据集按照 7:2:1 的比例划分为训练集、测试集和验证集,为模型训练、调参及性能评估提供可靠的数据基础,从而确保实验结果的可靠性和可重复性。

3.2 实验环境

本实验平台基于 Ubuntu 22.04 操作系统下进行, GPU 选择 NVIDIA GeForce RTX4090 (24 GB) 搭载 Intel

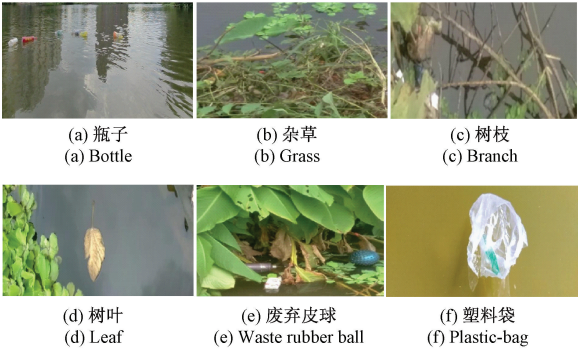


图 7 垃圾种类
Fig. 7 Types of garbage

Xeon Platinum 8352V 处理器, CUDA12.6, 深度学习框架为 PyTorch2.7.0, 编程语言环境为 Python3.10.16。训练关键参数如表 1 所示。

3.3 评价指标

实验在性能评估环节选取精准度(precision,P)、召回

表 1 模型训练关键参数

Table 1 Key parameters for model training	
参数名称	设置值
优化器	SGD
输入图片大小	640×640
学习率	0.001
训练轮数	250
批处理大小	24

率(recall,R)、每秒 10 亿次浮点运算数(GFLOPs),平均精度(mean average precision, mAP)来综合评估模型的性能,参数量 Params 来衡量模型的复杂程度。相关计算公式如式(20)~(23)所示。

$$P = \frac{TP}{FP + TP} \tag{20}$$

$$R = \frac{TP}{FN + TP} \tag{21}$$

$$AP = \int_0^1 P(r)dr \tag{22}$$

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \tag{23}$$

式中: TP 表示被正确识别为正类的样本数, FP 表示被错误判定为正类的样本数, FN 表示被错误判定为负类的样本数, n 为类别总数, k 为检测次数, AP 表示每种类别的平均准确率, mAP 表示所有类别 AP 的平均值, $mAP50$ 则为 IoU 阈值为 0.5 时的 mAP 。

3.4 FPSC_SCSA 模块引入不同注意力机制对比

为验证剔除 SPPF 最大池化层并引入不同膨胀率的共享卷积层的改进效果,同时评估多种注意力机制在特征提取过程中的不同效果,本文在 SPPF 和 FPSC 模块中分别嵌入性能优越的注意力机制,如坐标注意力(coordinate attention, CA)^[25]、压缩与激励(squeeze-and-excitation, SE)^[26]、卷积注意力模块(convolution block attention module, CBAM)^[27]等,并通过对比实验,系统分析不同机制对模型检测性能的影响,实验结果如表 2 所示。

由表 2 可知,在 SPPF 模块相同位置引入不同注意力机制,如 CBAM、SE、CA 和 SCSA 均能在一定程度上提升了检测性能。其中 SCSA 改进效果最为突出,相较于基线 mAP50 提高了 0.5%,计算量仅增加了 0.03 GFLOPs,展现了 SCSA 在多语义层次和通道特征的优化整合方面的优势。相比之下,引入 CBAM 注意力机制后参数量虽下降 0.04 M,但 mAP50 却下降 0.2%,这与预期的改进目标相悖。而引入 SE 和 CA 注意力机制虽均带来一定程度的精度提升,但提升幅度与 SCSA 相比分别低了 0.3% 和 0.1%,且计算量也比 SCSA 高了 0.01 和 0.03 GFLOPs,整体性价比较低。

FPSC 模块通过去除 SPPF 中的最大池化层,并引入不同膨胀率的共享卷积层,对原有 SPPF 结构进行重构。

表 2 空间金字塔池化模型对比实验

Table 2 Comparative experiments of spatial pyramid pooling models			
算法	mAP50/%	Params/10 ⁶	GFLOPs
SPPF	76.9	6.4	2.59
SPPF-CBAM	76.5	6.5	2.61
SPPF-SE	77.1	6.4	2.63
SPPF-CA	77.3	6.5	2.65
SPPF-SCSA	77.4	6.4	2.62
FPSC	77.6	6.4	2.64
FPSC-CBAM	77.2	6.5	2.66
FPSC-SE	77.9	6.4	2.68
FPSC-CA	78.3	6.5	2.71
FPSC-SCSA	78.8	6.4	2.67

实验结果表明,FPSC 的 mAP50 整体优于在 SPPF 模块中引入 CBAM、SE、CA 和 SCSA 的表现,FPSC 的 mAP50 相较基线提高了 0.7%,充分体现了多膨胀率卷积在捕捉尺寸差异较大特征方面的优势。其次,将 CBAM、SE、CA 和 SCSA 注意力机制引入 FPSC 模块,检测精度普遍高于其在 SPPF 模块中的表现,其中以 FPSC 与 SCSA 结合后的效果最佳。FPSC-SCSA 的计算量在较 FPSC 增加 0.03 GFLOPs 的情况下,mAP50 提升了 1.2%。相较基线 mAP50 更是提高了 1.9%,验证了本文提出的 FPSC_SCSA 模块对提升水面垃圾检测性能方面的有效性。

3.5 损失函数对比实验

为评估所提出的 SGSV-IoU 损失函数对模型性能的影响,本文选取几种性能优异的损失函数(CIoU、WIoU (weighted intersection over union)^[28]、EIoU (efficient intersection over union)^[29]、SIoU (scalable intersection over union)^[30]和 Shape-IoU)作为对比对象,将其分别应用于 YOLOv11n 基础模型中训练,并对比分析不同损失函数在水面垃圾检测任务中的检测效果,实验结果如表 3 所示。

表 3 损失函数对比实验

Table 3 Comparative experiments of loss functions %			
损失函数	P	R	mAP50
CIoU	74.2	72.9	76.9
WIoU	75.5	75.5	77.4
EIoU	77.5	76.3	77.7
SIoU	75.6	75.1	77.9
Shape-IoU	76.2	76.1	78.3
SGSV-IoU	76.5	76.4	78.7

根据表 3 所示,引入不同损失函数(CIoU、WIoU、EIoU、SIoU 和 Shape-IoU)后,模型的检测性能均得到一定程度的提升。其中,SIoU 在精确率方面优于其他损失函数,相较基线 mAP50 高了 0.6%,但在召回率和整体精度上表现一般。综合对比结果表明,Shape-IoU 效果最优,其 mAP50 较基线提升 1.4%,精确率和召回率分别提高 2%和 3.2%,表明 Shape-IoU 在小目标检测及长宽比例差异明显的目标定位任务中具备优势,有效克服了 CIoU 在边界细节刻画上的不足。

为验证改进效果,本文在 Shape-IoU 损失函数基础上融合 GSL 与 Varifocal Loss,构建 SGSV-IoU 损失函数,进一步提升模型的边界建模与目标分类能力,并与原损失函

数进行对比,验证改进的有效性。结果表明,SGSV-IoU 相较基准模型在精确度、召回率和 mAP50 上分别提升了 2.3%、3.5%、1.8%。与传统 Shape-IoU 相比,其精确度、召回率和 mAP50 分别提升 0.3%、0.3%和 0.4%。上述结果表明,所提出的 SGSV-IoU 损失函数在水面垃圾检测任务中整体表现优于其他损失函数,有效验证了 SGSV-IoU 损失函数的有效性。

3.6 消融实验

为验证本文所提出各改进模块的有效性及其相互兼容性,在自建的水面垃圾数据集上开展多组消融实验,实验结果如表 4 所示。表 4 中的“√”表示模型启用了相对应的改进模块。

表 4 消融实验
Table 4 Ablation experiments

组别	FCM-C3k2	FPSC-SCSA	SGSV-IoU	BIFPN-V2S	P/%	R/%	mAP50/%	Params/10 ⁶	GFLOPs
基线					74.2	72.9	76.9	2.59	6.4
1	√				76.3	75.5	79.1	2.68	6.4
2		√			75.5	75.4	78.8	2.67	6.4
3			√		76.5	76.4	78.7	2.59	6.4
4				√	76.4	76.2	78.2	1.75	6.2
5	√	√			78.5	75.3	79.2	2.52	6.3
6	√	√	√		81.2	76.8	79.4	2.58	6.3
7	√	√	√	√	80.1	77.3	79.7	1.78	6.3

由前 4 组实验结果可见,任一改进模块单独应用于基准模型上均能使其 mAP50 得到不同程度提升。在 4 组单一模块改进实验中,将主干网络后 3 个 C3k2 模块替换为 FCM-C3k2 模块的改进效果最为显著,mAP50 较基线提升 2.2%。在参数量与计算量方面,引入 BIFPN-V2S 模块的优化效果最为突出,分别减少了 0.84 M 和 0.2 GFLOPs,验证了其在减少模型参数冗余方面的优势。在准确率和召回率表现上,采用 SGSV-IoU 损失函数的改进效果最佳,较基线模型而言准确率和召回率分别提高了 2.3%和 3.5%。说明其在提高边界框回归精度方面具有显著优势。FPSC-SCSA 较基线而言在 GFLOPs 虽未降低,但在准确率、召回率和 mAP50% 分别提高了 1.3%、2.5%和 1.9%,体现了其能最大限度地减少关键信息的丢失的优势。

在验证完单个改进模块作用后,通过后续 3 组实验进一步考察各改进模块组合后对模型性能的影响,实验 5 将 FCM-C3k2 模块和聚合金字塔共享模块 FPSC-SCSA 同时引入模型中,较基线模型而言 mAP50 提升 2.3%,比单独引入 FCM-C3k2 模块在 mAP50 上提高了 0.1%,同时参数量和计算量分别减少 0.16 M 和 0.1 GFLOPs。实验 6 是在实验 5 的基础上加入损失函数 SGSV-IoU,较基线模型而言 mAP50 提高了 2.5%,且参数量和计算量也得到了有一定程

度的降低,与实验 5 模型相比虽然参数量上升了 0.06 M,但是准确率、召回率和 mAP50 上分别上升了 2.7%、1.5%和 0.2%,说明 SGSV-IoU 在提升边界框回归性能方面效果显著。实验 7 综合引入全部改进模块,相较基线模型准确率、召回率和 mAP50 分别提高了 5.9%、4.4%和 2.8%,参数量和计算量分别降低 0.81 M 和 0.1 GFLOPs。综上所述,实验 7 展现出最优的综合性能,进一步验证了算法改进的有效性。

3.7 主流算法对比实验

为进一步验证所提改进算法的优越性,将改进算法在自建的水面垃圾检测数据集上进行实验验证,并将实验结果与主流算法进行对比分析。对比算法包括双阶段(Faster_RCNN、Mask_RCNN)与单阶段(YOLOv3-ting、YOLOv5n、YOLOv6n、YOLOv8n、YOLOv9-ting、YOLOv10n、YOLOv11n),实验结果如表 5 所示。

实验数据表明,所提改进算法在检测精度上优于双阶段算法 Faster_RCNN 和 Mask_RCNN,mAP50 分别提升 4.5%和 3.4%,且改进算法的参数量与计算量远低于双阶段算法,在实时性方面具有显著优势。与单阶段 YOLO 算法相比,改进算法在 mAP50 上分别提升了 5.1%、4.3%、4%、3.8%、3.5%、3.1%和 2.8%。在模型参数方面,改进算法的参数量分别减少了 7.75、0.4、2.37、0.9、1.15、0.91

表 5 对比实验

Table 5 Comparative experiments

算法	P/%	R/%	F1%	mAP50/%	Params/ 10^6	GFLOPs
Faster_RCNN	73.8	72.1	71	75.2	41.3	206.7
Mask_RCNN	74.1	72.5	73	76.3	44.5	260.1
YOLOv3-tiny	72.1	71.6	71	74.6	9.53	14.3
YOLOv5n	73.1	71.4	71	75.4	2.18	5.8
YOLOv6n	73.2	71.6	71	75.7	4.15	11.5
YOLOv8n	74.1	73.2	72	75.9	2.68	6.8
YOLOv9-tiny	73.8	71.5	72	76.2	2.93	8.3
YOLOv10n	74.1	72.4	73	76.6	2.69	6.7
YOLOv11n	74.2	72.9	74	76.9	2.59	6.4
本文	80.1	77.3	76	79.7	1.78	6.3

和 0.81 M。在计算量方面,仅有 YOLOv5n 略低于改进算法,其余算法明显高于改进算法。综上,本文所提的改进算法在检测精度和轻量化上均优于主流目标检测算法,充分验证了所提算法的优越性。

3.8 实验结果可视化分析

为更直观地展示改进算法相较于 YOLOv11n 在水面垃圾检测任务中的性能优势,本文在自建数据集上进行了对比实验,并对不同类型水面垃圾的检测结果进行

了可视化分析,如图 8 所示。第 1 行为原始图像,第 2 和第 3 行分别对应 YOLOv11n 算法与本文改进算法的检测结果。由图 8 可知,YOLOv11n 在面对单目标、多目标及存在遮挡的目标时,均表现出检测精度不足的问题,尤其是在复杂背景下易出现漏检与误检现象。本文算法通过结构优化与特征增强,有效提升了模型对小目标与遮挡目标的识别能力,整体检测效果更为准确与稳定。

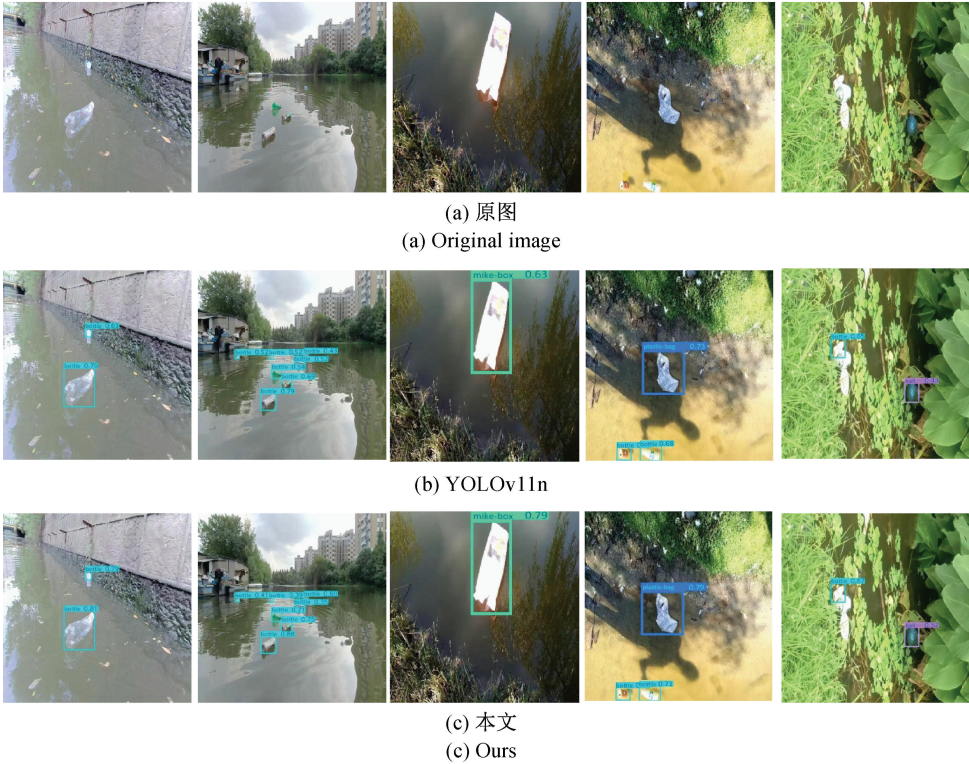


图 8 检测效果对比

Fig. 8 Detection results comparison

为更直观的展示本文所提模型的检测性能,绘制了相应的对比图,如图 9 所示,横坐标为训练轮数,纵坐标模型

的平均精度,红线为改进模型,蓝线为基准模型。由图 9 可知,改进模型在整体收敛速度明显优于基准模型,训练

前期(约40~90轮)检测精度略低于基准模型,这主要是由于改进后的聚合金字塔共享模块结构更为复杂,需要额外迭代以调整各层权重。但在训练达到90轮后,改进模型精度开始超过基准模型,并在250轮左右达到收敛,检测效果显著。上述实验结果充分验证了改进算法在检测精度和收敛性方面的优势,体现了其在水面垃圾检测任务中的应用价值。

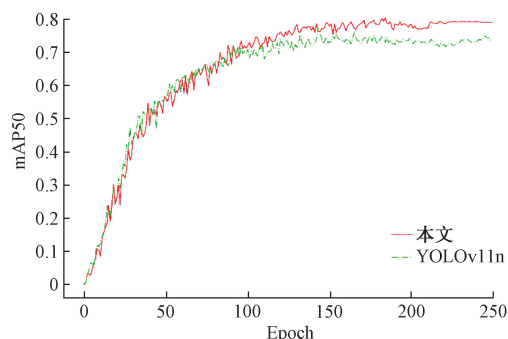


图9 平均检测精度对比实验

Fig. 9 Comparison of average detection accuracy

4 结 论

本文针对水面垃圾检测中目标易受水波干扰、光线反射及漂浮物遮挡等复杂背景影响而导致检测精度下降的问题,提出了一种改进的YOLOv11n检测算法。通过设计C3k2-FCM模块作为主干网络的部分特征提取模块,增强了模型对多尺度细节特征的提取能力;引入FPSC-SCSA模块,使模型能聚焦多个关键区域,减少关键信息的损失,提升模型对目标特征的区分能力。结合多尺度特征融合网络,构建BiFPN-V2S模块,实现了在不增加额外参数开销的前提下扩展感受野,强化全局信息交互能力,减少网络参数冗余,进一步提升特征提取性能;最后,构建SGSV-IoU损失函数,增强模型对边界形状和细节的刻画能力,提高对不规则漂浮物的定位精度。实验结果表明,改进后的算法在自建的水面垃圾数据集上较YOLOv11n基础模型,mAP50提高了2.8%,计算量降低了0.1 GFOPs,参数量减少了0.81 M,验证了改进模型的有效性与优越性。尽管模型性能已有所提升,但在计算复杂度方面仍有改进空间,后续研究将从参数优化与结构精简两方面入手,以在保证检测精度的同时进一步减低计算开销。

参考文献

[1] 安梦军, 吴朝明, 邓承志. MEC-YOLOv11n: 水面小目标漂浮物的检测算法[J]. 电子测量技术, 2025, 48(14): 185-197.
AN M J, WU ZH M, DENG CH ZH. MEC-YOLOv11n: Detection algorithm for small floating objects on water surface[J]. Electronic Measurement Technology, 2025, 48(14): 185-197.

[2] REN SH 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.
[3] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]. IEEE International Conference on Computer Vision, 2017: 2961-2969.
[4] DIWAN T, ANIRUDH G, TEMBHURNE J V. Object detection using YOLO: Challenges, architectural successors, datasets and applications[J]. Multimedia Tools and Applications, 2022, 82(6): 31-33.
[5] FENG CH J, ZHONG Y J, GAO Y, et al. TOOD: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2021: 3490-3499.
[6] BAI J, YANG S J. Detection method of marine floating garbage based on improved Faster R-CNN [C]. 3rd International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2022), 2022: 121-128.
[7] MA Y L, CHU ZH X, LIU H, et al. Strong and weak supervision combined with CLIP for water surface garbage detection[J]. Water, 2023, 15(17): 3156.
[8] LI X L, TIAN M J, KONG S H, et al. A modified YOLOv3 detection method for vision-based water surface garbage capture robot[J]. International Journal of Advanced Robotic Systems, 2020, 17(3), DOI: 10.1177/1729881420932715.
[9] DOHERTY J, GARDINER B, KERR E, et al. BiFPN-YOLO: One-stage object detection integrating bi-directional feature pyramid networks [J]. Pattern Recognition, 2025, 160: 112209.
[10] 石露露, 廖光忠. 改进YOLOv5s的明渠漂浮垃圾实时检测方法[J]. 计算机技术与发展, 2023, 33(9): 83-90.
SHI L L, LIAO G ZH. Real-time detection method for floating debris in open channels based on improved YOLOv5s[J]. Computer Technology and Development, 2023, 33(9): 83-90.
[11] 冀越文, 张涛. 面向无人艇的轻量级水面小目标检测算法[J]. 全球定位系统, 2025, 50(4): 20-30.
JI Y W, ZHANG T. Lightweight detection algorithm for small water-surface target on unmanned surface vehicles[J]. Global Positioning System, 2025, 50(4): 20-30.
[12] SI Y ZH, XU H Y, ZHU X ZH, et al. SCSA:

- Exploring the synergistic effects between spatial and channel attention[J]. *Neurocomputing*, 2025, 634: 129866.
- [13] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection [C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020: 10778-10787.
- [14] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[J]. *ArXiv preprint arXiv:1803.01534*, 2018.
- [15] CELAYA A, RIVIERE B, FUENTES D. Generalized surface loss for reducing the Hausdorff distance in medical imaging [J]. *ArXiv preprint arXiv:2302.03868*, 2023.
- [16] ZHANG H Y, WANG Y, DAYOUB F, et al. VarifocalNet: An IoU-aware dense object detector[J]. *ArXiv preprint arXiv:2008.13367*, 2020.
- [17] ZHANG H, ZHANG S J. Shape-IoU: More accurate metric considering bounding box shape and scale[J]. *ArXiv preprint arXiv:2312.17663*, 2023.
- [18] XIAO Y, XU T F, XIN Y, et al. FBRT-YOLO: Faster and better for real-time aerial image detection[J]. *ArXiv preprint arXiv:2504.20670*, 2025.
- [19] PENG Y P, SONKA M, CHEN D Z. U-Net V2: Rethinking the skip connections of U-Net for medical image segmentation[J]. *ArXiv preprint arXiv:2311.17791*, 2024.
- [20] 童小钟, 魏俊宇, 苏绍璟, 等. 融合注意力和多尺度特征的典型水面小目标检测[J]. *仪器仪表学报*, 2023, 44(1): 212-222.
- TONG X ZH, WEI J Y, SU SH J, et al. Typical small water-surface target detection based on attention fusion and multi-scale features[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(1): 212-222.
- [21] 徐尽达, 陈慈发, 张上. 基于轻量级算法的水上垃圾小目标检测研究[J]. *电子测量技术*, 2024, 47(18): 145-154.
- XU J D, CHEN C F, ZHANG SH. Research on small water-surface debris detection based on lightweight algorithms[J]. *Electronic Measurement Technology*, 2024, 47(18): 145-154.
- [22] 徐守坤, 龚研, 李宁. FA-YOLOv7: 水面垃圾小目标检测算法[J]. *计算机工程与设计*, 2025, 46(8): 2328-2334.
- XU SH K, GONG Y, LI N. FA-YOLOv7: Small target detection algorithm for water surface debris[J]. *Computer Engineering and Design*, 2025, 46(8): 2328-2334.
- [23] 张瑶, 陈姚节. 改进 YOLOv8 的水面小目标检测算法[J]. *计算机系统应用*, 2024, 33(4): 152-161.
- ZHANG Y, CHEN Y J. Improved YOLOv8-based detection algorithm for small water-surface targets[J]. *Computer Systems & Applications*, 2024, 33(4): 152-161.
- [24] 程亮, 杨渊, 张云飞, 等. 面向无人艇智能感知的水上目标识别算法研究[J]. *电子测量与仪器学报*, 2021, 35(9): 99-104.
- CHENG L, YANG Y, ZHANG Y F, et al. Research on water surface target recognition algorithm for unmanned surface vehicle intelligent perception[J]. *Journal of Electronic Measurement and Instrumentation*, 2021, 35(9): 99-104.
- [25] KIM J H, AN Y J, KIM J. Improving speech emotion recognition through focus and calibration attention mechanisms[J]. *ArXiv preprint arXiv:2208.10491*, 2022.
- [26] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 42(8): 2011-2023.
- [27] WOO S Y, PAEK J CH, LEE J Y, et al. CBAM: Convolutional block attention module [C]. *European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 3-19.
- [28] ZHANG Y F, REN W Q, ZHANG ZH, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146-157.
- [29] PAN H Y, YU S Q. A systematic IoU-related method: beyond simplified regression for better localization[J]. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 2021, DOI:10.1109/TIP.2021.3077144.
- [30] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression [J]. *ArXiv preprint arXiv:2205.12740*, 2022.

作者简介

王海群, 硕士, 副教授, 研究生导师, 主要研究方向为智能控制与应用、深度学习。

E-mail: wanghq0604@163.com

谢伟民(通信作者), 硕士研究生, 主要研究方向为深度学习、目标检测。

E-mail: 2029208690@qq.com

晁帅, 博士, 主要研究方向为深度学习、目标检测。

于海峰, 博士, 主要研究方向为深度学习、图像处理。