

赵聪钊. 基于改进 YOLOv8 模型的水面漂浮物检测算法[J]. 智能计算机与应用, 2025, 15(7): 111–117. DOI: 10.20169/j.issn.2095-2163.250716

基于改进 YOLOv8 模型的水面漂浮物检测算法

赵聪钊

(河北工程大学 信息与电气工程学院, 河北 邯郸 056000)

摘要: 水面漂浮物目标检测是保护水域生态环境的重要任务,光照和水质条件的影响会导致漏检率高、目标类别不准确等问题。为解决以上问题,本文提出了一种改进 YOLOv8 模型的水面漂浮物检测算法。首先,引入了感受野注意力机制(RFA),加强不同尺度特征提取能力;其次,使用分布移位卷积 DSConv 替换了 C2f 层的标准卷积,提高模型的运算速度;最后,使用了损失函数 Slide loss,让模型更加关注难识别的物体。实验结果表明,改进后的算法精度 $mAP@0.5$ 较 YOLOv8 原网络提升了 3.3%,达到了 85.1%,参数量增长不到 1%。改进后算法具备较强的通用性和泛化能力,适用于不同光照和水质条件下的监测。

关键词: 水面漂浮物; 目标检测; YOLOv8; 感受野注意力; DSConv; Slide loss

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2025)07-0111-07

Floating object detection algorithm based on improved YOLOv8 model

ZHAO Congzhao

(School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056000, Hebei, China)

Abstract: Floating object detection on water surfaces is a crucial task for preserving the ecological environment of water bodies. Challenges such as variations in lighting conditions and water quality can result in issues like high false-negative rates and inaccurate target categorization. To address these challenges, an improved algorithm for detecting floating objects on water surfaces based on YOLOv8 is proposed in this paper. Firstly, the algorithm incorporates the Receptive-Field Attention mechanism to enhance the capability of extracting features at different scales. Subsequently, the standard convolution in the C2f layer is replaced with the DSConv to boost the computational speed of the model. Finally, the Slide loss function is employed to make the model pay more attention to challenging-to-identify objects. Experimental results demonstrate that the improved algorithm achieves a 3.3% increase in accuracy ($mAP@0.5$) compared to the original YOLOv8 network, reaching 85.1%, with a parameter growth of less than 1%. The enhanced algorithm exhibits strong generality and generalization capabilities, making it suitable for monitoring under varying lighting and water quality conditions.

Key words: floating; object detection; YOLOv8; Receptive-Field Attention; DSConv; Slide loss

0 引言

随着社会经济的飞速发展,水体污染问题日益严峻,海洋、湖泊、河流等各类水体的水面漂浮物越来越多。水面垃圾大量堆积,不仅污染水质,还对水体生态平衡产生负面影响,威胁着水生生物的生存,这也破坏了生态景观,甚至可能影响河道和水坝的正常运行,对居民用水安全构成潜在威胁^[1]。

水面漂浮物包括塑料、海绵、玻璃和金属等,传统的监测方法往往依赖于人工观察或传感器网络,受限于成本、时效性以及受到水面环境复杂性的影响。在

图像处理早期,王斌^[2]采用局部峰值奇异性进行图像预处理和基于亮度区域划分的目标检测机制,但受到水面环境多变和目标众多的干扰;汤伟等^[3]结合 Meanshift 算法和最大类间方差改进的色差灰度模型,虽然提升了精度,但实时性不高,不适合移动硬件部署。杨鹏^[4]使用机器学习方法实现了对水面漂浮物的实时识别,利用均值漂移算法将采集到的水面漂浮物图像分割成若干个小块,并对水面污染程度进行大致估算,对漂浮物的颜色矩量特征分别提取 RGB 色彩空间和 HIS 色彩空间,采用支持向量机实现了水面漂浮物的分类;於双飞等^[5]利用 HSV 空间代替

作者简介: 赵聪钊(2001—),男,硕士研究生,主要研究方向:计算机视觉。Email: zcz1020@126.com。

收稿日期: 2023-11-28

哈尔滨工业大学主办 ◆ 专题设计与应用

RGB 空间完成对原画面不同色彩范围阈值分割,提出了一种可以减少水面波纹影响的漂浮物检测方法。总体而言,传统的检测方法普遍存在检测速度慢、精度低和鲁棒性差的问题。

随着深度学习技术的不断进步,针对水面漂浮物的目标检测问题有了新的解决方案。与传统机器学习相比,基于深度学习的目标检测算法利用多层神经网络直接对图像执行卷积和池化操作,从而直接提取图像的本质特征。在图像识别领域,深度学习表现出显著的优势^[6]。雷李义等^[7]提出了一个包含水草和落叶的小型漂浮物数据集,在预训练好的 Faster-RCNN (Faster Region - based Convolutional Neural Network) 和 SSD (Single Shot MultiBox Detector) 等模型上进行微调,取得了一定的识别效果,表明 CNN (Convolutional Neural Network) 可用于水面漂浮物识别,但是这些模型的网络结构相对复杂。随着 YOLO (You Only Look Once) 模型的出现,徐浩^[8]、李德鑫^[9]、李国进等^[10]对 YOLO 模型进行改进并成功应用于水面目标检测,改进后的模型较之前的模型精度较高,模型较小。尽管目前深度学习在漂浮物检测方面取得了一定的进步,但仍存在一些问题,如对某些物体的识别精度较低,模型参数量仍较大等。

针对受光照和水质条件的影响,导致算法漏检率高、目标定位不准确、目标类别不准确等问题,本文引入 YOLOv8 模型,结合其先进的目标检测特性,并对 YOLOv8 模型进行了改进。

1 YOLOv8 模型

YOLOv8 是 Ultralytics 基于 YOLOv5 在 2023 年

推出的一种单阶段目标检测模型,继承了先进的 SOTA (State-Of-The-Art) 技术,并引入了全新的方法并改进,以进一步提高模型的性能和灵活性。YOLOv8 共分为 YOLOv8-n、s、m、l、x 5 个模型,这些模型的参数量和计算量随着深度和宽度的增大而增加,使用者可以根据自身应用场景选择适合的网络结构。例如,YOLOv8n 在保持相对较快检测速度和较高准确度,适用于嵌入式系统和低成本设备。YOLOv8 算法模型主要由 3 个部分构成:

(1) 主干特征提取模块 (Backbone): 该模块采用了 CSPDarkNet (Cross Stage Partial DarkNet) 结构,用于从输入图像中提取特征;C2f 模块替代了先前 YOLO 模型的 C3 模块,采用了梯度流更加丰富的结构,以提高模型的性能。

(2) 特征加强模块 (Neck): 该模块采用了 PANet (Path Aggregation Network) 结构,通过上采样和通道融合,将 PANet 的 3 个输出分支送入检测模块,可以对从主干模块提取的特征进行进一步处理,以增强模型对目标的表达能力。

(3) 检测模块 (Head): 用解耦头结构进行检测,将回归分支和预测分支分离,有助于加速模型的收敛;该模块负责生成目标检测的结果,包括目标的位置、类别等信息。

YOLOv8 的这些模块协同工作,通过优化主干特征提取模块、调整通道数以及引入特征加强模块和解耦头结构的检测模块,提高模型在目标检测任务上的性能,包括收敛速度和效果,在多个计算机视觉任务中表现出色。YOLOv8 算法模型的基本结构如图 1 所示。

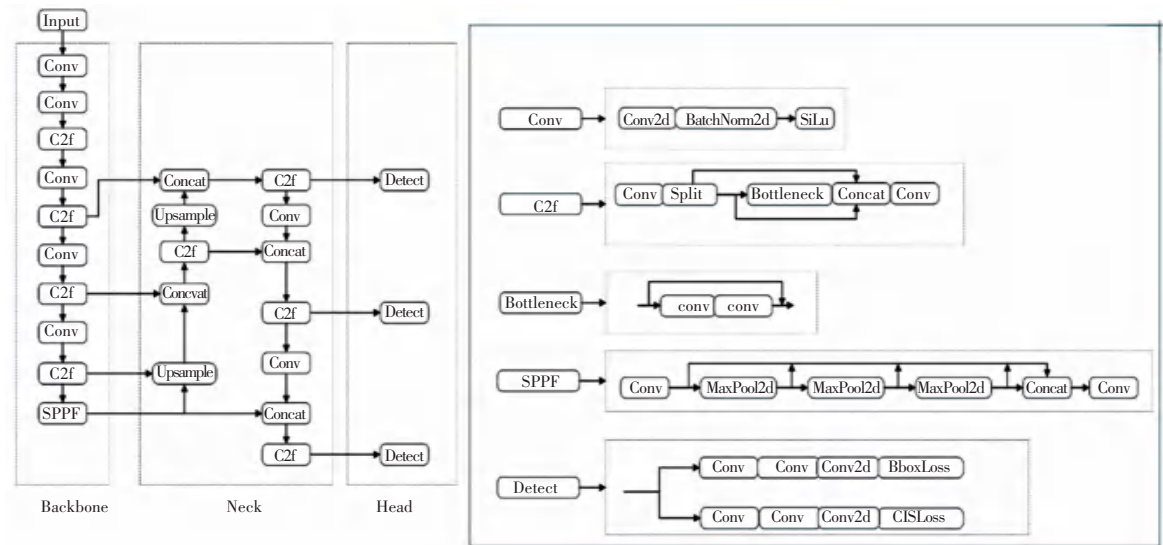


图 1 YOLOv8 算法模型结构
Fig. 1 YOLOv8 algorithm model structure

2 改进的 YOLOv8 模型

针对受光照和水质条件的影响导致算法漏检率高、目标定位不准确等问题,本文对 YOLOv8 模型进行了改进:

(1) 在 Conv 模块中加入感受野注意力模块 RFA(Receptive-Field Attention),可以充分利用特征图信息,加强不同尺度特征提取能力,RFA 不仅关注感受野空间特征,还为大尺寸卷积核提供有效的注意力权重,解决水面受到水质条件和光照的影响

导致漂浮物难以识别的问题;

(2)水面漂浮物中目标排列不规则,将 C2f 模块中的 Conv 使用 DSConv (Distribution Shift Convolution)代替以适应水面漂浮物中目标物体的不同角度,加快模型的运算速度;

(3)由于水面漂浮物中目标的形状和尺寸多种多样,选择使用损失函数 Slide loss,将更多注意力集中在难分类的错误示例上。改进后 YOLOv8 算法模型结构如图 2 所示。

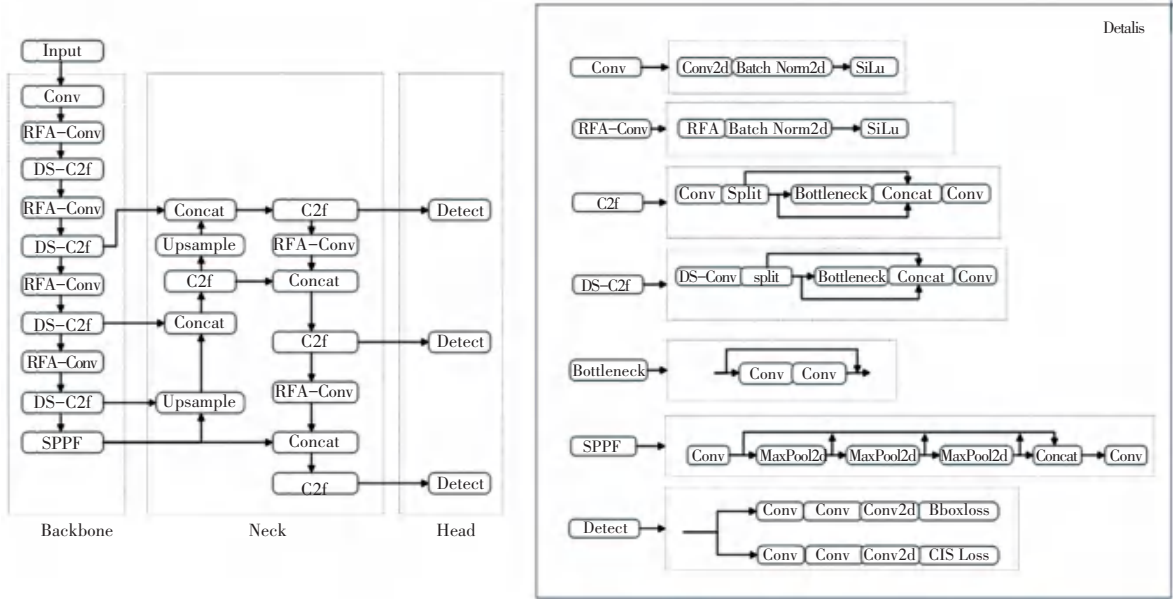


图 2 改进后 YOLOv8 算法模型结构

Fig. 2 Improved YOLOv8 algorithm model structure

2.1 感受野注意力

尽管空间注意力在提高卷积神经网络性能方面取得了显著成就,但仍然存在一些局限性,特别是在处理大尺寸卷积核时,空间注意力生成的注意力图中包含的信息相对有限。为了解决这一问题,引入了感受野注意力 RFA。RFA 的核心思想是将注意力从空间特征转移到感受野空间特征中。现有的空间注意力机制,如卷积块注意力模块(CBAM)和协调注意力(CA),虽然关注了空间特征,但没有考虑整个感受野的空间特征,并没有完全解决卷积核参数共享的问题,无法有效解决大规模卷积核的参数共享问题^[11-12]。因此,空间注意机制的有效性是有限的,而 RFA 不仅关注感受野空间特征,而且为大尺寸卷积核提供了有效的注意力权重。

为了将 RFA 引入卷积神经网络中,本文使用感受野注意卷积运算(RFAConv),这是一种替代标准卷积运算的新方法,相比于标准卷积运算,RFAConv

在计算成本和参数增量上几乎可以忽略不计,同时显著提高了网络性能。3×3 尺寸卷积核的 RFAConv 的整体结构如图 3 所示。

图 3 给出通过展开方法提取 3×3 接收场空间特征的示例。设输入 $X \in RC \times H \times W$,为了方便显示,将步长设置为 3,展开后维数变为 $9C \times H \times W$, C 、 H 和 W 分别表示通道数、输入的高度和宽度。滑窗展开操作用于提取感受野空间特征,将输入的图像切分成多个局部区域,并将其展开成一个矩阵形式,为后续卷积运算准备数据。然而滑窗展开的计算速度相对较慢,所以在 RFAConv 中,利用一种快速方法来提取接收场空间特征,当使用 3×3 卷积核提取特征时,接收场空间特征中的每个 3×3 大小的窗口代表一个接收场滑块,使用快速组卷积提取接收场特征后,原始特征被映射成新的特征,展开方法更快、更有效。

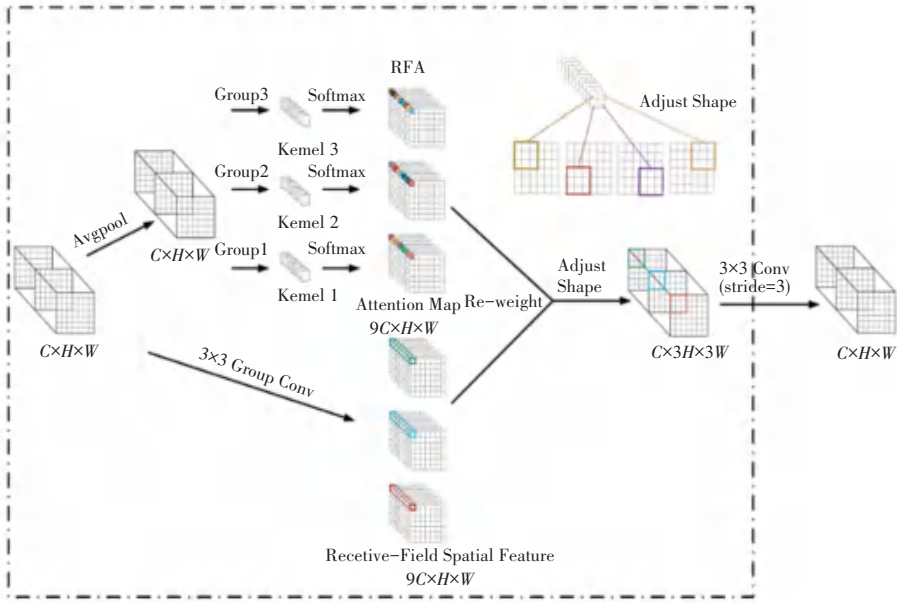


图 3 3×3 尺寸卷积核的 RFACnv 整体结构

Fig. 3 Overall structure of RFACnv with 3×3 convolutional kernels

RFACnv 中的交互接收场特征信息学习注意力图可以增强网络性能,但与每个接收场特征交互可能导致额外的计算开销。为了最小化计算开销和参数数量,利用平均池化(AvgPool)来聚合每个接收场特征的全局信息;使用 1×1 组卷积运算来交互信息;最后,使用归一化指数函数(Softmax)来强调接收场特征中每个特征的重要性。通常, F 是通过将注意力图 A_{rf} 与变换后的感受场空间特征 F_{rf} 相乘而得, RFA 的计算可以表示为:

$$F = \text{Softmax}(g^{1 \times 1}(\text{AvgPool}(X))) \times \text{ReLU}(\text{Norm}(g^{k \times k}(X))) = A_{rf} \times F_{rf} \quad (1)$$

其中, F 代表通过 RFA 操作得到的特征图,其中考虑了感受野空间特征,并通过注意力机制强调了不同特征的重要性; $g^{k \times k}$ 表示大小为 $k \times k$ 的分组卷积; k 表示卷积核的大小; Norm 表示归一化; X 表示输入特征图。

2.2 分布移位卷积 DSConv

分布移位卷积 DSConv 是一种卷积层变体,其可以替换到标准神经网络架构中,并实现更低的内存使用和更高的计算速度,可以提高标准卷积的存储效率和速度,降低模型的运算量。

DSConv 的框架图如图 4 所示。从图 4 可知, DSConv 是使用量化和分布位移来模拟卷积行为的方法。DSConv 由两个部分组成:可变量化内核(Variable Quantized Kernel, VQK)和分布位移。VQK 只保存位长可变的整数值,作为 DSConv 量化组件的一部分,有助于提高乘法速度并提升存储效率。分布

位移的目的是调整 VQK 的分布,以模拟原始卷积核的分布,“位移”包括缩放和偏置操作,通过两个张量实现即内核分布移位器(Kernel Distribution Shifter, KDS)和通道分布移位器(Channel Distribution Shifter, CDS)。KDS 负责移动 VQK 每个块中的分布,而 CDS 则负责移动每个通道中的分布。

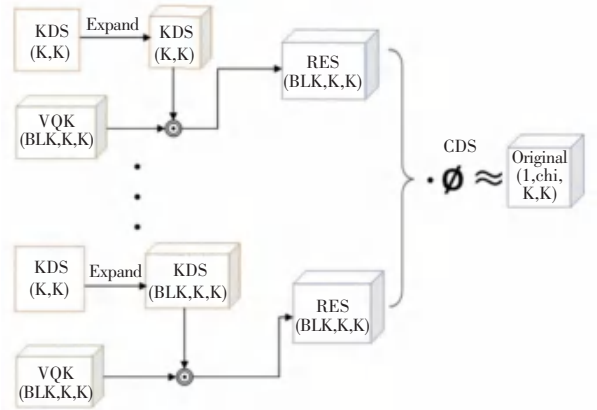


图 4 DSconv 框架图

Fig. 4 DSconv framework diagram

通过 DSConv,卷积核的大小可以减少到原始大小的一部分,实现更快、更节省内存的计算;通过量化和位移操作的结合,在保持相对较小的计算和内存开销的同时,能够模拟原始卷积操作的效果。预训练网络时,将权重张量滤波器按深度划分为长度为 B 的块,所有的块共享一个浮点值,然后对这些块进行量化处理。每个张量的内存节约可通过下式表示:

$$p = \frac{b}{32} + \frac{\sum_{i=1}^B C_i}{C_i} \quad (2)$$

其中 C_i 代表通道数, b 代表所选择的超参数设置。

2.3 滑动损失 Slide Loss

损失函数是衡量模型预测与实际目标之间差异的函数,用于衡量模型的性能并指导模型参数的更新,以便逐步提高模型的预测准确性。为了提高水面漂浮物中难以分类物体的精度,在 YOLOv8 模型中选择使用滑动损失函数 Slide Loss,可以更灵活的更新模型参数,该损失函数将所有边界框的 IoU (Intersection over Union) 值的平均值作为阈值 μ , 小于 μ 的作为负样本,大于 μ 的作为正样本。但靠近边界的样本往往分类不明确而遭受较大的损失,为了模型能够学会优化这些样本,并更充分地利用来训练网络,该损失函数通过强调边界处的样本,为困难的样本分配更高的权重。Slide Loss 损失函数的具体形式如下:

$$f(x) = \begin{cases} 1, & x \leq \mu - 0.1 \\ e^{1-\mu}, & \mu - 0.1 < x < \mu \\ e^{1-x}, & x \geq \mu \end{cases} \quad (3)$$

Slide Loss 损失函数如图 5 所示,其自适应地学习正样本阈值参数和负样本阈值参数。在 μ 附近设置较高的权重会增加难分类示例的相对损失,从而将更多注意力集中在难分类的错误示例上。Slide Loss 损失函数可以提升目标检测算法的性能,实现更准确的水面漂浮物定位和分类。

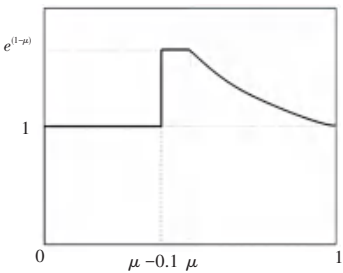


图 5 Slide Loss 损失函数
Fig. 5 Slide Loss function

3 实验与结果

3.1 数据集和评价指标

数据集由网络数据集整合后扩充而来,数据集包含各种光照条件和水质的图片,共有 6 648 张图片,分辨率为 1 024×1 024。将漂浮物分为 8 类:塑料、海绵、金属、泡沫、玻璃、网、浮标和绳子,按照 7 : 2 : 1 的比例分为训练集、验证集和测试集。

mAP (平均精度)表示所有标签的平均精确率之和除以总类别数, mAP 的值越大,则说明模型的

平均准确率越高。模型性能评估中, IoU 表示预测框与真实标注框之间的重叠程度,通常使用不同的 IoU 阈值来确定预测结果的正确与否,当 IoU 大于等于设定阈值时,认为预测正确;当 IoU 小于设定阈值时,认为预测错误。当 mAP 的 IoU 阈值设置为 0.5 时表示为 $mAP@0.5$,当 IoU 阈值设置为 0.5 ~ 0.95 时表示为 $mAP@0.5 \sim 0.95$ 。

模型参数量 (Params) 表示模型训练过程中的参数总量,包括权重、偏置值等,用来衡量模型的空间复杂度和规模,是评估模型轻量化的重要指标。 FPS (Frames Per Second) 表示每秒处理的图像帧数,是衡量模型训练速度和性能的重要指标之一。

本次实验通过 $mAP@0.5$ 、 $mAP@0.5 \sim 0.95$ 、参数量和 FPS 这 4 个评价标准来评估目标检测模型的性能。

3.2 实验环境

模型训练环境为 Windows 11 操作系统,CPU 为 Intel(R)Core(TM) i9-12900F,内存为 16 GB,显卡为 Nvidia GeForce RTX 3080Ti,显存为 12 G,深度学习框架为 Pytorch-GPU 2.1.0+cu121。训练参数设置:训练周期 (epochs) 为 600,批处理量 (batch size) 为 32,进程数 (workers) 为 8,输入图像尺寸为 1 280×1 280。

3.3 实验结果与分析

为验证本文改进的 YOLOv8n 模型的轻量化和出色的检测性能,将改进后的模型与现阶段主流的目标检测模型: Faster R - CNN^[13]、YOLOv3^[14]、YOLOv5s、YOLOv7tiny^[15] 和 YOLOv8n 进行了对比实验,实验结果见表 1。可见本文改进后模型的精度高于对比模型,分别提升了 18.9%、12.6%、5.8%、4.2% 和 3.3%,而且本文模型参数量仅为 3.0 M, FPS 也为较高水平,契合了水面漂浮物需求中轻量化、高精度的特点,适合部署在低成本、低算力、计算资源有限设备上。

表 1 各网络模型实验结果对比

Table 1 Comparison of experimental results across different networks

模型	Params/ M	$mAP@0.5$ / %	$mAP@0.5 \sim 0.95$ / %	FPS (f/s)
Faster R-CNN	42	66.2	56.3	13
YOLOv3	12.2	72.5	61.1	21
YOLOv5s	7	80.3	70.5	84
YOLOv7tiny	6	80.9	70.9	91
YOLOv8n	3	81.8	71.2	105
本文模型	3	85.1	74.1	102

为验证各个模块改进后的性能,又进行了一系列的消融实验,实验结果见表 2,表中√表示使用该改进方法。实验结果表明,最终改进后算法的精度

均高于其它改进方法的搭配,表明本文算法的改进均有效,改进后的算法参数增长量仅为 0.02 M,适合水面漂浮物目标检测。

表 2 YOLOv8n 消融实验
Table 2 YOLOv8n ablation experiment

Method	RFA	DSconv	Slide Loss	Params/M	$mAP@0.5/\%$	$mAP@0.5 \sim 0.95/\%$
YOLOv8n				3.01	81.8	71.2
A	√			3.03	83.7	72.9
B		√		3.01	82.5	71.8
C			√	3.01	82.9	72.1
D	√	√		3.03	84.3	73.5
E	√		√	3.03	84.6	73.7
F		√	√	3.01	83.3	72.5
G	√	√	√	3.03	85.1	74.1

3.4 检测效果与分析

原图与检测结果对比如图 6 所示,可以发现

YOLOv8 出现了误检、错检和漏检的情况,本文模型的检测效果明显优于 YOLOv8 模型。

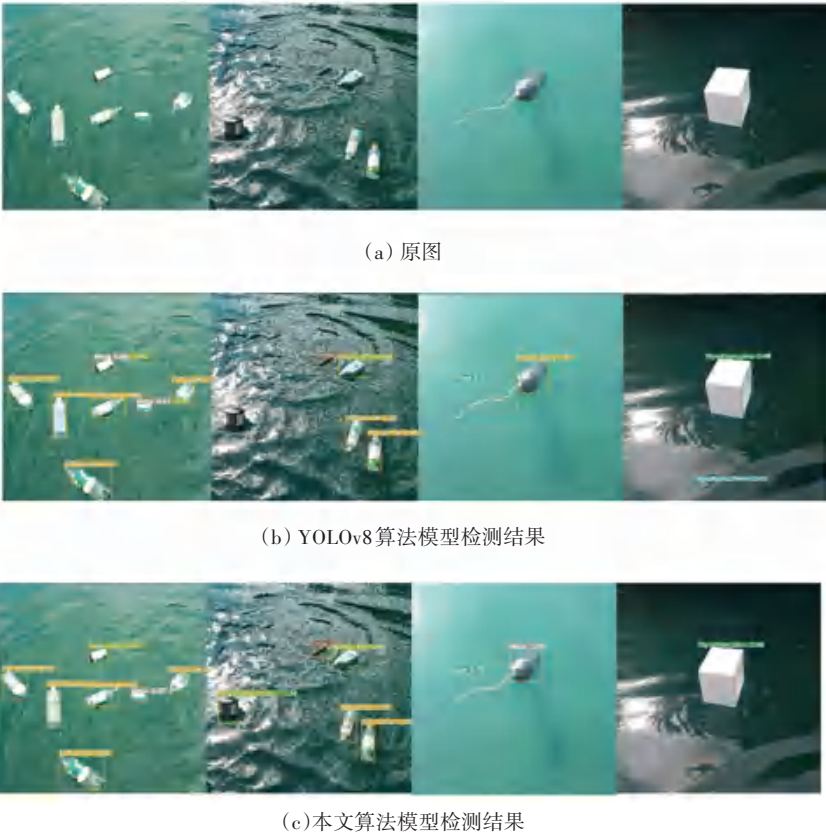


图 6 原图与算法检测结果图

Fig. 6 Original image and algorithm detection result image

4 结束语

本文对水面漂浮物目标检测问题进行了深入研究,在 YOLOv8 模型基础上引入 RFA 感受野注意力机制、DSConv 卷积和 Slide Loss 损失函数,提高了算

法的检测准确性, $mAP@0.5$ 达到了 85.1%,解决了受光照和水质条件的影响会导致目标定位不准确、目标类别不准确等问题,同时在模型参数量上的微小增长也确保了算法的轻量级性能。实验结果表明,改进后的算法相对于原始 YOLOv8 具有更高的

检测性能,适合在资源受限的设备上进行部署。本文改进后的算法模型有望为生态污染监测和环保执法提供可靠的技术支持,为改善水域生态系统的健康状况提供更为可行的解决方案。

未来的研究将致力于进一步提升算法的适应性,以应对更为复杂和多变的实际场景。

参考文献

[1] 褚晶,王进. 水面漂浮物主动清除装置[J]. 今日制造与升级, 2022,154(12):81-84.

[2] 王斌. 海面目标红外检测方法研究[D]. 大连: 大连海事大学, 2018.

[3] 汤伟,刘思洋,陈景霞,等. 基于计算机视觉的水面目标分割与识别方法[C]//2018 中国自动化大会(CAC2018). 2018;532-537.

[4] 杨鹏. 基于计算机视觉的水面智能监控研究[D]. 贵阳: 贵州民族大学,2015.

[5] 於双飞,杨怀斌,孔凡国,等. 一种水面漂浮目标的视觉检测方法[J]. 机电工程技术, 2019, 48(4):145-147.

[6] 耿创,宋品德,曹立佳. YOLO 算法在目标检测中的研究进展[J]. 兵器装备工程学报,2022,43(9):162-173.

[7] 雷李义,艾娇燕,彭婧,等. 基于深度学习的水面漂浮物目标检测评估[D]. 南宁:广西大学,2019.

[8] 徐浩. 基于深度学习的水面漂浮物目标识别算法研究[D]. 上海: 华东师范大学,2022.

[9] 李德鑫,闫志刚,孙久运. 基于无人机视觉的河道漂浮垃圾分类检测技术研究[J]. 金属矿山,2021,543(9):199-205.

[10] 李国进,姚冬宜,艾娇燕,等. 基于改进 YOLOv3 算法的水面漂浮物检测方法[J]. 广西大学学报(自然科学版),2021,46(6): 1569-1578.

[11] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018; 3-19.

[12] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway,NJ;IEEE,2021: 13713-13722.

[13] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.

[14] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv,1804.02767, 2018.

[15] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 7464-7475.