

改进 YOLOv8 的水面小目标检测算法^①



张 瑶¹, 陈姚节^{1,2}

¹(武汉科技大学 计算机科学与技术学院, 武汉 430081)

²(智能信息处理与实时工业系统湖北省重点实验室, 武汉 430081)

通信作者: 陈姚节, E-mail: chenyaojie@wust.edu.cn

摘 要: 针对水面目标检测中的噪声干扰和小目标的漏检问题, 提出一种改进 YOLOv8 的水面小目标检测算法 YOLOv8-WSSOD (YOLOv8-water surface small object detection). 首先, 为降低水面复杂环境在主干网络下采样过程中产生的噪声干扰, 提出基于 BiFormer 双层路由注意力机制构建的 C2fBF (C2f-BiFormer) 模块, 在特征提取过程中保留细粒度的上下文特征信息; 其次, 针对水面小目标的漏检问题, 新增一个更小的检测头, 提升网络对小目标的感知力, 并在 Neck 端引入 GSConv 和 Slim-neck, 减轻模型复杂度并保持精度; 最后, 使用 MPDIoU 损失函数解决 CIoU 损失函数的局限性, 以提高模型检测准确率. 实验结果表明, 相较于原始 YOLOv8 算法, 该算法在水面小目标上平均准确率 $mAP@0.5$ 提升了 4.6%, $mAP@0.5:0.95$ 提升了 2.2%, 并且改进后的算法检测速度达到 86f/s, 能有效实现对水面小目标快速、准确的检测.

关键词: YOLOv8; 水面小目标检测; BiFormer; GSConv; MPDIoU

引用格式: 张瑶, 陈姚节. 改进 YOLOv8 的水面小目标检测算法. 计算机系统应用, 2024, 33(4): 152–161. <http://www.c-s-a.org.cn/1003-3254/9445.html>

Improved YOLOv8 Algorithm for Small Object Detection on Water Surface

ZHANG Yao¹, CHEN Yao-Jie^{1,2}

¹(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China)

²(Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430081, China)

Abstract: To address the problems of noise interference and missed detection of small objects in water surface object detection, this study proposes an improved You Only Look Once version 8 (YOLOv8) algorithm for water surface small object detection, namely, YOLOv8-WSSOD. Specifically, to reduce the noise interference caused by the complex water surface environment during the downsampling in the backbone network, the study proposes the C2f-BiFormer (C2fBF) module constructed based on BiFormer's bi-level routing attention mechanism to retain fine-grained contextual feature information during feature extraction. Then, as to the missed detection of small objects on the water surface, a smaller detection head is added to enhance the network's sensitivity to small objects. At the Neck end, the ghost-shuffle convolution (GSConv) and Slim-neck structures are used to reduce the model's complexity and maintain precision. Finally, the limitations of the complete intersection over union (CIoU) loss function are overcome by the minimum point distance-based IoU (MPDIoU) loss function to improve the model's detection precision. The experimental results show that compared with the original YOLOv8 algorithm, the proposed algorithm increases the mean average precision $mAP@0.5$ and $mAP@0.5:0.95$ on small objects on the water surface by 4.6% and 2.2%, respectively. Furthermore, the modified algorithm, achieving a detection speed of 86 f/s, is readily available for fast and accurate detection of small objects on the water surface.

Key words: YOLOv8; small object detection on water surface; BiFormer; GSConv; MPDIoU

① 基金项目: 装备发展部“慧眼行动”项目 (62602010214)

收稿时间: 2023-09-24; 修改时间: 2023-10-20; 采用时间: 2023-11-09; csa 在线出版时间: 2024-03-04

CNKI 网络首发时间: 2024-03-08

目前,水面漂浮物检测已成为智能无人船领域的研究热点.作为无人船路径规划的首要任务,水面漂浮物检测过程面临各种环境噪声干扰,如天气、光线变化、水面反射等,而随着人工智能的发展,研究人员针对这些问题提出了许多基于深度学习的检测算法.这些算法大致可以分为两类:两阶段算法和单阶段算法.常用的两阶段算法包括:R-CNN、Faster R-CNN、Mask R-CNN等.Zhang等^[1]通过融合浅层特征和深层特征,提出了一种改进Faster R-CNN的水面目标检测方法,确保了检测多尺度水面目标时的鲁棒性和准确性;刘伟等^[2]为及时发现并处理水面漂浮垃圾,采用不同的特征提取网络,提出了一种基于Mask R-CNN模型的水面漂浮物识别算法,提高了水面目标检测的平均准确率和平均召回率.

另一种是单阶段算法,不产生候选框区域,通过回归直接预测物体的类别概率和位置信息.目标检测中,YOLO (you only look once) 系列算法和SSD (single shot multibox detector) 算法是常用的单阶段算法.尽管两阶段算法在检测精度上略微优于单阶段算法,但在整体上,其检测速度相较于单阶段算法明显不足.因此,在实际应用中,单阶段算法受到的关注度更高.李国进等^[3]基于改进YOLOv3的水面漂浮物检测算法,通过增强K-means聚类和在预测支路中添加类别激活映射改善了传统卷积网络对水面目标检测精度低、定位不准等问题;王林等^[4]利用深度可分离卷积和自适应空间特征融合等模块改进YOLOv4网络,实现了对海上船舶目标的高精度实时检测;Deng等^[5]在YOLOv5主干网络的每个CSP单元中添加ECA注意力机制,并对特征融合模块做轻量化改进,提出一种基于无人船水面检测场景的ATT-YOLOv5-Ghost算法,解决了检测算法参数量大、检测速度慢和梯度计算重复问题;陈任飞等^[6]采用图像降噪和增强算法处理原始图像,并以MobileNetV2作为特征提取网络,提出一种改进的SSD算法,提高了水面漂浮物的检测效率;张晓鹏等^[7]结合暗通道去雾算法、SE注意力机制以及改进非极大值抑制,提出一种基于YOLOv5的海上船舶识别模型(SE-NMS-YOLOv5),解决了多目标和雾天条件下,海上船舶检测准确率低的问题.然而,这些方法并未考虑到水面小目标.目标检测中将目标所占像素点数与原图总像素点数之比小于0.12%或者尺寸小于32×32像素的目标定义为小目标.小目标的图像尺寸较小,具

有分辨率低、覆盖像素较少、定位精度差和特征表达能力弱等特点.因此在复杂的水面背景下小目标会更容易受到外界因素的干扰,进而导致其在实际检测中经常出现漏检.

YOLOv8是YOLO系列的基本模型之一,相较于原来的anchor-based检测方法,它所采用的anchor-free方法具有更高的检测精度和检测速度.但在复杂水面环境下,由于小目标漂浮物特征复杂、背景多样,使用YOLOv8检测算法仍存在着定位误差和对目标感知能力不足等缺点.因此,为解决上述问题,本文针对YOLOv8的改进提出YOLOv8-WSSOD算法,首先为了减轻主干网络下采样过程中的噪声影响,引入捕获远程依赖思想,采用BiFormer^[8]双层路由注意力机制构建C2fBF模块,保留特征提取过程中更细粒度的上下文信息;其次,针对小目标漏检问题,添加一个更小的检测头提升网络对小目标的感知力,并在Neck端引入GSConv和Slim-neck^[9]技术保持精度并降低计算量;最后使用MPDIoU^[10]损失函数替换CIoU损失函数,提高模型的泛化能力和精度.

1 相关工作

1.1 YOLOv8 网络模型

如图1, YOLOv8算法模型主要包含4个模块:输入端、主干网络、Neck端和输出端.输入端对输入网络的图片通过Mosaic数据增强、自适应图片缩放和灰度填充等方式进行预处理.主干网络中,Conv、C2f和SPPF结构通过卷积、池化等方式提取图片特征.Neck端基于PAN (path aggregation network) 结构设计,通过上采样、下采样和特征拼接对不同缩放尺度特征图进行融合.输出端采用decoupled head结构,实现分类和回归过程的解耦,包含正负样本匹配和损失计算.YOLOv8网络采用TaskAlignedAssigner^[11]方法,对分类分数和回归分数加权,并以加权结果匹配正样本.损失计算包含分类损失和回归损失计算,采用BCE (binary cross-entropy) 计算分类损失,使用DFL (distribution focal loss)^[12]和CIoU (complete intersection over union) 损失函数计算回归损失.

1.2 Transformer 简介

近年来,Transformer^[13]不仅在自然语言处理领域取得可观成果,同时在计算机视觉领域也获取了巨大成功.当前,基于Transformer设计实现的模型有很多,如图像

分类模型的 ViT^[14], 目标检测模型的 DETR^[15]. 然而在小目标检测研究方向, 远距离小目标像素点少, CNN 的下采样机制导致提取的特征缺乏充足的上下文信息. 相比 CNN, Transformer 的自注意力机制在小目标检测中能够有效地聚合全局信息, 捕获远距离依赖关系, 将对

象之间的信息与位置感知有效关联. Transformer 通过自注意力机制使得模型可以对输入数据的不同部分进行加权处理, 从而可以捕获到远程依赖关系, 但自注意力机制需要对输入的每个元素与其他所有输入元素进行交互计算, 会产生巨大的计算负担和沉重的内存占用.

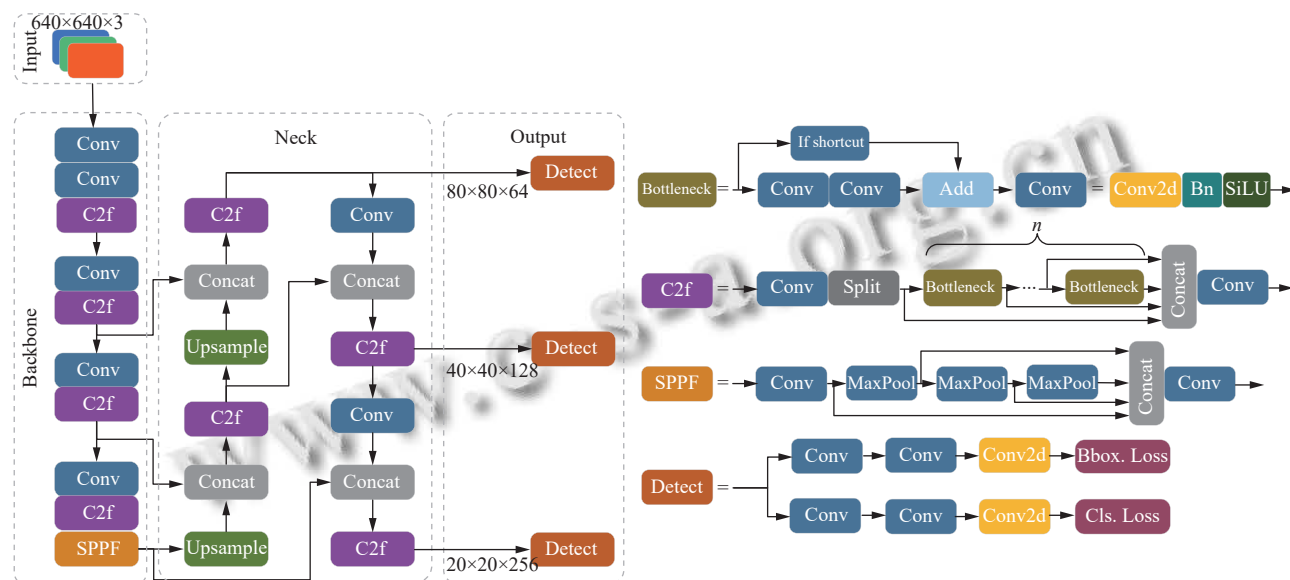


图1 YOLOv8 网络结构

2 本文算法

本文选择 YOLOv8n 作为基线模型, YOLOv8n 模型参数量低, 且检测速度快, 适用于本文算法研究. 本文在此基础上提出的 YOLOv8-WSSOD 算法, 其网络如图 2 所示, 与 YOLOv8 算法网络结构图相比, 本文的 YOLOv8-WSSOD 算法在主干部分使用提出的 C2fBF 模块替换第 1 个 C2f 模块, 在 Neck 端通过上采样、特征拼接等方式获得一个 160×160 大小的浅层特征图以对应更小的检测头, 另外使用 GSConv 替换原来的 Conv 模块、VoVGSCSPC 模块替换原来的 C2f 模块.

2.1 主干网络改进

在复杂水面环境中图像会因为天气、光照变化以及水面的反射等产生噪声, 故可能存在以下几种问题.

- (1) 强光照射使得水面图像出现剧烈的反射, 如图 3(a) 所示.
- (2) 拍摄环境光线较暗使得图像模糊失真, 如图 3(b) 所示.
- (3) 雨、雾等天气会对检测目标产生遮挡, 以及图像中存在密集水面目标时, 物体间出现重叠, 如图 3(c)、图 3(d) 所示.

卷积神经网络 (CNN) 在提取这些场景图像特征的过程中, 可能会产生大量噪声. 这些噪声将会破坏像素之间的远程依赖关系, 减弱模型的检测与识别能力, 从而可能导致误检甚至漏检.

本文为了降低噪声对模型的干扰, 引入捕获远程依赖思想, 以此来提高水面目标检测与识别效果. BiFormer 是基于 Transformer 模型的一个变体, 它通过双层路由注意力机制 (bi-level routing attention) 实现了更灵活的计算分配和特征感知.

$$A^r = Q^r (K^r)^T \quad (1)$$

$$I^r = \text{topkIndex}(A^r) \quad (2)$$

$$K^g = \text{gather}(K, I^r), V^g = \text{gather}(V, I^r) \quad (3)$$

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \quad (4)$$

如图 4, 双层路由注意力机制首先将输入特征图 $X \in \mathbb{R}^{H \times W \times C}$ 划分成 $S \times S$ 个非重叠区域, 使得 X 转化为 $X^r \in \mathbb{R}^{S^2 \times HWC/S^2}$, 然后采用全连接层对 X^r 进行线性投影, 得到 Q 、 K 和 V , 并计算 Q 、 K 的均值分别得到 Q^r 、 K^r , 再通过式 (1) 计算邻接矩阵 A^r , 衡量不同区域之间的语义相似度. 式 (2) 对矩阵 A^r 进行过滤, 从中挑

选出语义相似度最高的前 k 个区域, 得到 I' , 其中 I' 的第 i 行包含与第 i 个区域最相关的 k 个区域的索引. 式 (3) 利用 I' 对 K 和 V 进行过滤和收集, 得到 K^s 和 V^s . 最后,

对 Q 、 K^s 和 V^s 集中应用细粒度的 token-to-token 注意力, 如式 (4) 所示, 其中 LCE 是深度可分离卷积, 卷积核大小为 5, 步长为 1.

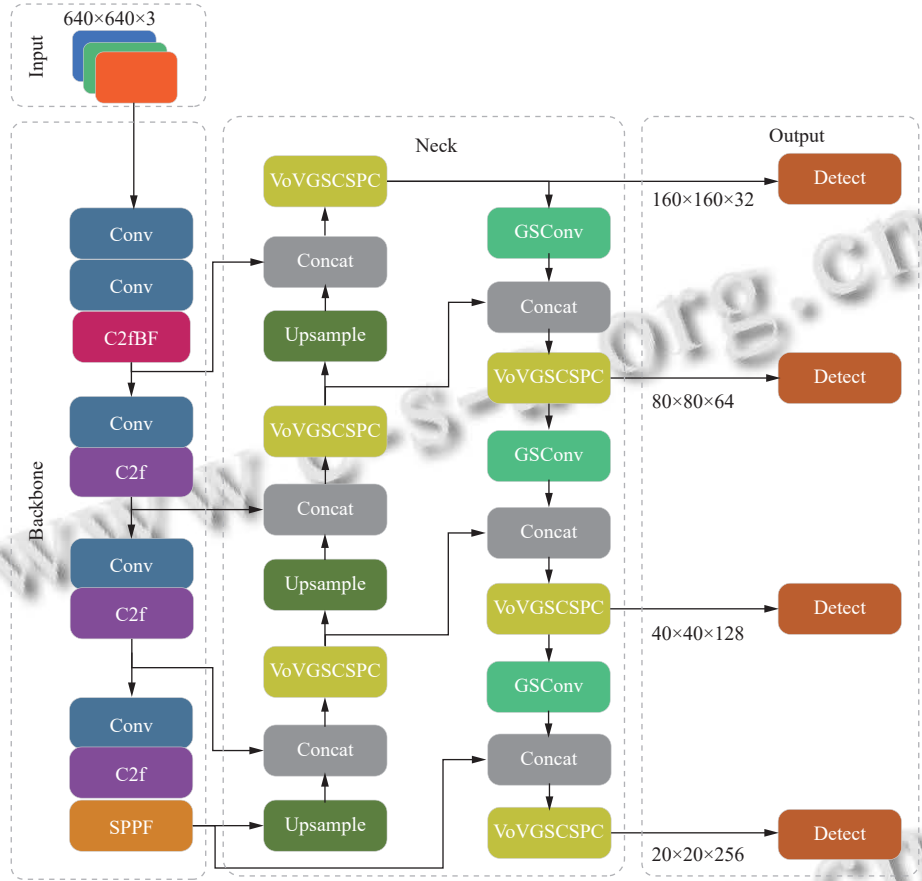


图2 YOLOv8-WSSOD 网络结构



图3 水面噪声干扰

双层路由注意力机制在粗糙区域过滤掉大部分不相关的键值对, 去除冗余信息, 只保留小部分路由区域, 并在这部分路由区域再应用注意力机制, 因此比直接在全局使用自注意力机制可以降低计算量. BiFormer

结合双层路由注意力机制构建的 BiFormerBlock 结构如图 5(a) 所示, 在开始时使用深度可分离卷积 (DWConv) 隐式编码输入的相对位置信息, 然后依次使用双层路由注意力机制和多层感知机 (MLP) 模块对输入的信息进行跨位置关系建模和逐位置嵌入.

上述模块中, 一方面, 使用 BiFormerBlock 对输入特征图实现动态稀疏关注, 不仅能降低噪声干扰, 而且与一般的 Transformer 自注意力结构相比计算量更小, 极大地减轻了内存负担; 另一方面, C2f 拥有更多的分支跨层连接, 能丰富目标检测模型的梯度流, 实现对残差特征的充分学习. BiFormerBlock 模块和 C2f 模块均能提升模型目标检测任务的精度. 因此, 针对水面小目标检测中存在的噪声问题, 本文结合 BiFormerBlock 和 C2f 模块的优点, 提出 C2fBF (C2f-BiFormer) 模块 (如图 5(b)), 用以替换主干网络的第 1 个 C2f 模块, 在

保持模型轻量的同时,使模型既能完成对残差特征的学习,也能最大程度上保留细粒度的上下文特征信息,

降低水面小目标检测中的各种噪声干扰影响,实现模型对远程依赖关系建模。

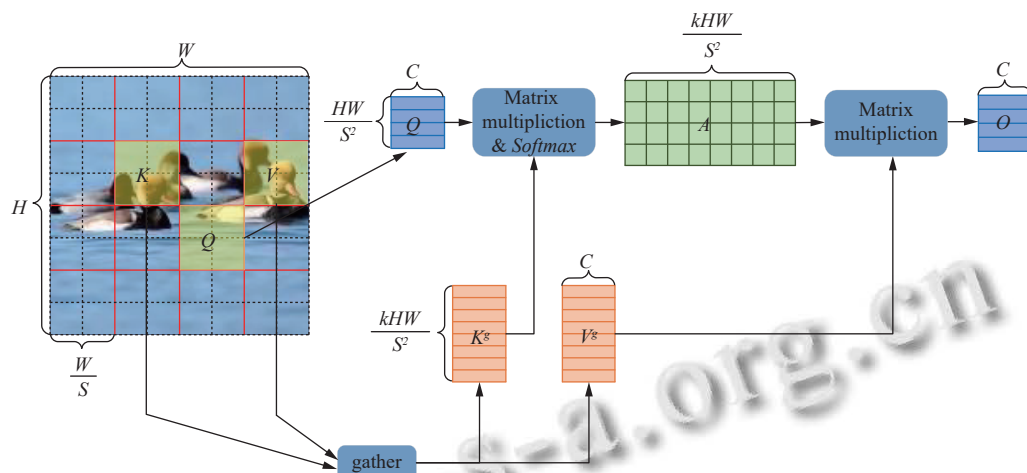


图4 双层路由注意力机制

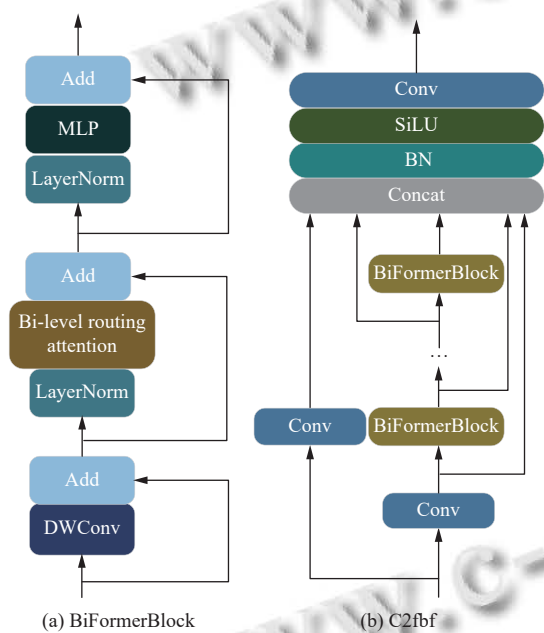


图5 BiFormerBlock、C2fBF 模块

2.2 Neck 端改进

原YOLOv8设定的3个检测头,对应的Neck端特征图大小是 80×80 、 40×40 、 20×20 ,分别负责检测小、中、大型尺度目标。然而,YOLOv8的检测头对小目标的检测能力仍然有限,且在主干网络和Neck端的浅层特征图中包含更丰富的小目标信息没有被充分利用,故在此基础上通过上采样和特征拼接得到一个大小 160×160 的特征图,用以对应更小的检测头,提高小目标的检测准确率。

由于添加特征图会导致Neck端网络层增多,造成

网络参数量和计算量上升,所以引入GSConv和Slim-neck技术,在保持精度基础上降低模型的复杂度和计算量。

如图6(a)所示,GSConv模块由标准卷积、深度可分离卷积组合构建,通过Shuffle混合操作,使得标准卷积生成的信息融入深度可分离卷积生成的信息中,以保证通道间的信息交互并降低时间损耗。

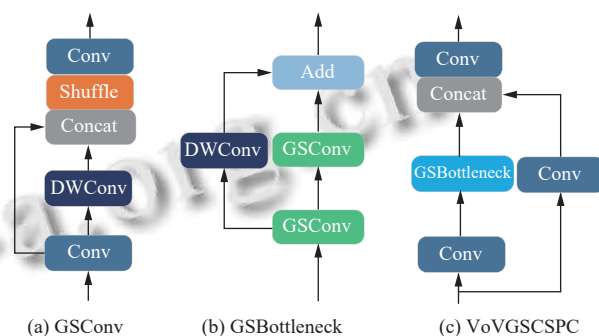


图6 GSConv, GSBottleneck, VoVGSCSPC 模块

如图6(b)所示,GSBottleneck是由两个GSConv模块和一个DWConv模块组成,输入的特征信息分别导入两个GSConv模块和一个DWConv模块,最后把这些模块的输出相加。在GSBottleneck基础上,VoVGSCSPC使用一次性聚合方式构建而成,降低网络计算量和参数量的同时,保证模型的精度,如图6(c)所示。

2.3 损失函数改进

CIoU作为YOLOv8网络中的边界框回归损失函数,通过预测框和真实框的重叠面积、中心点距离和长宽比来计算损失,公式如下:

$$\mathcal{L}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(B_{\text{gt}}, B_{\text{pred}})}{C^2} + \alpha v \quad (5)$$

$$\text{IoU} = \frac{B_{\text{gt}} \cap B_{\text{pred}}}{B_{\text{gt}} \cup B_{\text{pred}}} \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan \frac{w_{\text{pred}}}{h_{\text{pred}}} \right)^2 \quad (7)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (8)$$

其中, B_{gt} 表示真实框, B_{pred} 表示预测框; $\rho^2(B_{\text{gt}}, B_{\text{pred}})$ 代表预测框和真实框中心点的欧氏距离; C 表示能够包

含预测框和真实框的最小外接矩形的对角线长度; α 为平衡参数; v 用来衡量长宽比是否一致。

在水面小目标检测的边界框回归过程中, 当预测框与真实框发生中心点重叠时, 此时 CIoU 损失函数可以通过惩罚项对预测框与真实框之间的形状、位置和尺寸偏差进行优化, 但这种情况下当预测出的边界框与真实框长宽比一致而预测的长宽值不相等时 (如图 7 所示, 绿框是真实框, 红框是预测框), CIoU 损失函数会失去有效性, 此时不利于损失函数的充分表达, 将会限制模型的收敛速度和准确性。

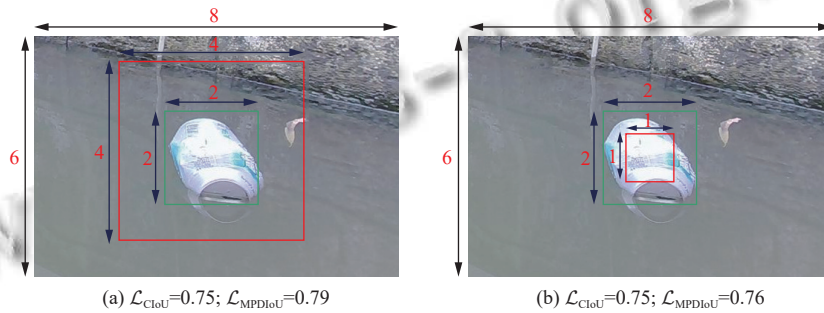


图 7 两种不同预测情况下的损失函数计算结果

因此, 采用 MPDIoU 损失函数替换原网络中的 CIoU 损失函数, 将最小点距离纳入考虑, 使用最小点距离度量对损失函数进行重新描述, 减少损失函数的总自由度, MPDIoU 损失函数的计算公式如式 (9) 所示, 其中 w 、 h 分别为网络中输入图片的宽和高. 式 (10) 中 $(x_1^{\text{pred}}, y_1^{\text{pred}})$ 、 $(x_2^{\text{pred}}, y_2^{\text{pred}})$ 分别是预测框的左上角和右下角坐标, 式 (11) 中 $(x_1^{\text{gt}}, y_1^{\text{gt}})$ 、 $(x_2^{\text{gt}}, y_2^{\text{gt}})$ 则分别是真实框的左上角与右下角坐标。

$$\mathcal{L}_{\text{MPDIoU}} = 1 - \text{IoU} + \frac{d_1^2}{h^2 + w^2} + \frac{d_2^2}{h^2 + w^2} \quad (9)$$

$$d_1^2 = (x_1^{\text{pred}} - x_1^{\text{gt}})^2 + (y_1^{\text{pred}} - y_1^{\text{gt}})^2 \quad (10)$$

$$d_2^2 = (x_2^{\text{pred}} - x_2^{\text{gt}})^2 + (y_2^{\text{pred}} - y_2^{\text{gt}})^2 \quad (11)$$

MPDIoU 损失函数考虑到了 CIoU 损失函数的局限性, 更适合用于水面小目标检测中. 根据 MPDIoU 损失函数的公式可知, 在中心点非重叠等其他情况下, 该函数能促进预测框向真实框靠近, 并且在预测框和真实框中心点重叠、长宽比一致但预测出的长宽值不相同, MPDIoU 函数的惩罚项不为 0, 不会退化为 IoU 损失. 因此, 相比 CIoU 损失函数, 使用 MPDIoU 损失

函数不仅简化了计算过程, 还能够稳定模型收敛, 提高模型对水面小目标的检测准确率。

3 实验与结果分析

3.1 实验环境和参数设置

实验在 Ubuntu 22.04 系统环境下进行, 硬件设备为 Intel i5-12400F 处理器、一块 12 GB 显存的 NVIDIA GeForce RTX 3060 显卡和 32 GB 内存的计算机, 软件环境为 Python 3.7, 使用 PyTorch 1.11.0 框架. 采用 YOLOv8n 作为基线网络模型, 设置超参数 learning rate 为 0.01, batch size 为 8, epoch 为 150.

3.2 实验数据集

本文数据集主要来源于内河漂浮垃圾数据集 Flow^[16]和水面目标检测数据集 WSODD^[17]. Flow 是欧卡智船发布的全球第 1 个真实内河场景下无人船视角的漂浮垃圾检测数据集, 其子数据集 Flow-Img 包含 2 000 张图片, 其中超过 1/2 的目标都是小目标 (Area < 32×32 像素); WSODD 是 Zhou 等^[17]针对水面自动驾驶提出的一个大规模、高质量的标准基准数据集, 共 7467 张水面图像, 包含 14 个类别和 21911 个标注实例。

本文从 WSODD 数据集中手工筛选出 3 000 张小

目标图片加入到 Flow-Img 子数据集中, 得到 5 000 张包含大量水面小目标的图片, 此外为了扩充数据集, 本文还从英特网上收集了 1 000 张水面小目标图片, 利用 LabelImg 对这 6 000 张图片重新进行 YOLO 格式的标注, 分成 4 个类别: 小船 (Boat)、浮标 (Buoy)、垃圾 (Rubbish)、动物 (Animal)。为满足实验需求, 本文以 6:2:2 比例将得到的 6 000 张数据集划分为训练集、验证集和测试集。

3.3 评价指标

本文采用平均精度均值 (mean average precision, mAP)、模型计算量 (giga floating-point operations per second, GFLOPs) 和每秒检测图片数量 (frames per second, FPS) 对模型进行评价。 mAP 用以评价模型的准确性, 其计算公式如下:

$$mAP = \sum P_A / N \quad (12)$$

其中, P_A 表示由精确率 (Precision) 作为横轴、召回率 (Recall) 作为纵轴构成的曲线围成的区域面积, N 代表检测类别的总数。 $mAP@0.5$ 表示 IoU 阈值设置为 0.5 时, 计算每一类的平均精度 AP (average precision), 然后对所有类别 AP 取的平均值。 $mAP@0.5:0.95$ 是指 IoU 阈值从 0.5 开始到 0.95, 以 0.05 为步长的每个 IoU 阈值计算一个 mAP , 并对最后的所有 mAP 求的平均值。

3.4 改进方法效果对比

为探索本文提出的改进方法实际效果, 分别对 YOLOv8n 主干网络、Neck 端和损失函数进行改进, 并优化调整参数, 通过实验对比各种改进方法的效果。

3.4.1 主干网络改进实验

YOLOv8n 模型的主干网络中共包含 4 个 C2f 模块, 为研究改进主干网络结构的合适位置, 对每个 C2f 模块进行替换, 采用提出的 C2fBF 模块, 并测试改进后模型的精度。

由表 1 可知, 引入 C2fBF 模块会导致模型计算量上涨, 同时 $mAP@0.5$ 最高也会提升 1.6%, 但在主干网结构中, 大量地使用 C2fBF 模块会影响检测准确率的提高, 使得提升的 $mAP@0.5$ 降低, 推测是因为网络层数的增加导致计算量上升的同时, 也加剧了数据流的阻力。根据实验结果, 本文将 YOLOv8n 主干网中第 1 个 C2f 模块替换为 C2fBF 模块。

3.4.2 Neck 端改进实验

相比于原始 YOLOv8n 的 Neck 端, 本文提出的

YOLOv8-WSSOD 算法在 Neck 端添加了一个 160×160 大小的浅层特征图以对应更小的检测头, 同时使用 GSConv 和 VoVGSCSPC 分别替换了原来的 Conv 和 C2f 模块, 以保持精度并降低计算量。

表 1 改进主干网络后的模型性能对比

Model	$mAP@0.5$ (%)	GFLOPs	Layers
YOLOv8n	77.8	8.2	225
YOLOv8n+1 C2fBF	79.4	9.1	244
YOLOv8n+2 C2fBF	78.8	10.1	277
YOLOv8n+3 C2fBF	78.4	11.0	310
YOLOv8n+4 C2fBF	77.9	11.7	329

实验结果如表 2 所示, 添加更小的检测头, 可以显著提高模型对小目标的检测准确性, $mAP@0.5$ 提高了 2%, $mAP@0.5:0.95$ 提高了 1.1%, 说明在 160×160 大小的特征图中包含了大量小目标的信息。本文还将 GSConv 与 Slim-neck 分别用在 YOLOv8n 和添加了小检测头的 YOLOv8n 模型上, 结果表明 GSConv 和 Slim-neck 不仅能够降低模型最高约 11% 的计算量, 还略微提高了模型检测准确率 (表 2)。

表 2 Neck 端改进后的模型性能对比

Model	$mAP@0.5$ (%)	$mAP@0.5:0.95$ (%)	GFLOPs
YOLOv8n	77.8	37	8.2
YOLOv8n+Smallerhead	79.8	38.1	12.7
YOLOv8n+GSConv & Slim-neck	78	37.1	7.4
YOLOv8n+Smallerhead+GSConv & Slim-neck	80.1	38.2	11.3

3.4.3 损失函数改进实验

将 MPDIoU 损失函数和常用的损失函数 IoU 、 $DIoU$ ^[18]、 $GIoU$ ^[19]、 $CIoU$ 进行实验对比, 由表 3 可得, 使用 MPDIoU 损失函数的 YOLOv8n 模型检测准确率更高, 相比使用 $CIoU$ 损失函数的原 YOLOv8n 模型, $mAP@0.5$ 提升 1.3%, $mAP@0.5:0.95$ 提升 0.6%, 说明使用 MPDIoU 损失函数能够使模型对边界框的回归更加稳定, 预测精度更高。

表 3 改进损失函数后的模型性能对比 (%)

Model	$mAP@0.5$	$mAP@0.5:0.95$
YOLOv8n+ IoU	76.9	36.6
YOLOv8n+ $DIoU$	77.4	36.9
YOLOv8n+ $GIoU$	77.2	36.7
YOLOv8n+ $CIoU$	77.8	37
YOLOv8n+MPDIoU	79.1	37.6

3.5 消融实验

从表 4 可见: 在主干网中将第 1 个 C2f 模块替换

为 C2fBF 模块, 水面小目标检测准确率 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提高了 1.6% 和 0.7%; 在特征融合模块添加了浅层特征图对应小目标检测头, 可以提升模型对小目标的感知力, $mAP@0.5$ 提升了 2%, $mAP@0.5:0.95$ 上升了 1.1%, 另外引入 GSConv 和 Slim-neck, 能使模型计算量下降 9.7%, 减轻计算负担; 使用 MPDIoU 损失函数提高了模型的检测精度, 有效克服了小目标

检测真实场景下 CIoU 损失函数的限制. 综合实验结果, 同时使用 C2fBF 模块、小目标检测头、GSConv 和 Slim-neck 以及 MPDIoU 损失函数, 经过 150 轮迭代训练, 最后训练得到 YOLOv8-WSSOD 模型比原 YOLOv8n 模型提高了 4.6% 的 $mAP@0.5$ 和 2.2% 的 $mAP@0.5:0.95$, 尽管检测速度 FPS 有所下降, 但仍然维持了 86 f/s 的较高速度 (表 4, 图 8).

表 4 消融实验结果

C2fBF	Smallerhead	GSConv & Slim-neck	MPDIoU	$mAP@0.5$ (%)	$mAP@0.5:0.95$ (%)	GFLOPs	FPS (f/s)
—	—	—	—	77.8	37	8.2	110
√	—	—	—	79.4	37.7	9.1	91
—	√	—	—	79.8	38.1	12.7	94
—	—	√	—	78	37.1	7.4	112
—	—	—	√	79.1	37.6	8.2	109
√	√	—	—	81.4	38.6	13.6	85
√	√	√	—	81.5	38.7	12.2	88
√	√	√	√	82.4	39.2	12.2	86

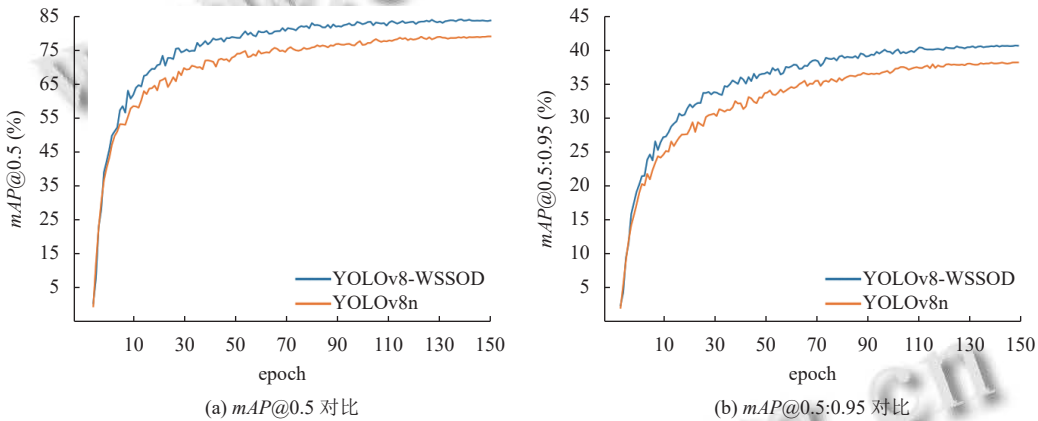


图 8 模型改进前后的 mAP 曲线对比

3.6 不同模型的检测对比实验

为验证本文算法 YOLOv8-WSSOD 在水面小目标检测方面的优越性, 采用相同的数据集、实验条件, 将 YOLOv8-WSSOD 与主流的目标检测算法, 包括 Faster R-CNN^[20]、SSD^[21]以及 YOLO 系列模型 YOLOv5、YOLOv7^[22]等进行测试对比, 同时为了深入比较, 本文还选择了针对水面目标检测改进的 ATT-YOLOv5-Ghost 算法和针对小目标检测改进的 TPH-YOLOv5^[23]算法进行对比.

从表 5 可得, 本文提出的算法对比 YOLOv5s、YOLOv7-tiny 和 YOLOv8n, $mAP@0.5$ 分别提高了 6.4%、9.5% 和 4.6%, $mAP@0.5:0.95$ 分别提高了 3.2%、4.8% 和 2.2%. 本文算法相比 YOLOv5s、YOLOv7-tiny 减少了计算量, 检测速度也分别提高了 2 f/s 和 8 f/s, 虽然相比 YOLOv8n, 本文算法检测速度有所下

降, 但是仍然满足实时检测要求. TPH-YOLOv5 在 YOLOv5s 的基础上增加了检测头和自注意力机制, 提高了模型对小目标的检测精度, 因此其对水面小目标检测效果较好, 但网络计算量较大, 且检测速度相较于 YOLOv5s 下降明显. 另外, 改进的 ATT-YOLOv5-Ghost 算法相比 YOLOv5s 提高了检测精度和检测速度, 降低了计算量, 而本文算法提出的改进方法对水面目标特征学习明显更为充分, 在 ATT-YOLOv5-Ghost 的基础上使得检测准确率 $mAP@0.5$ 提高了 5.3%, $mAP@0.5:0.95$ 提高了 2.6%. 最后本文改进的网络与 Faster R-CNN、SSD 网络相比, 计算量大幅下降, 检测速度和平均精度也都有较大提升. 由此可见, 改进后的 YOLOv8-WSSOD 算法在显著提高检测精度的同时, 保持了较高的检测速率, 综合性能优于目前主流检测算法和改进后的其他算法.

表5 不同模型的检测性能对比

Model	输入尺寸	$mAP@0.5$ (%)	$mAP@0.5:0.95$ (%)	FPS (f/s)	GFLOPs
Faster R-CNN	720×720	64.3	27.9	26	947.3
SSD	512×512	62.4	26.8	53	273.7
YOLOv5s	640×640	76.0	36.0	84	16.0
TPH-YOLOv5	640×640	81.6	38.7	69	29.9
ATT-YOLOv5-Ghost	640×640	77.1	36.6	94	14.5
YOLOv7-tiny	640×640	72.9	34.4	78	13.0
YOLOv8n	640×640	77.8	37.0	110	8.2
YOLOv8-WSSOD	640×640	82.4	39.2	86	12.2

3.7 可视化分析

为了验证本文改进的算法在实际水面目标检测不同背景下的有效性,从实验测试集中选取在内河、湖泊和海面背景下拍摄的图片进行测试,检测效果分别如图9(a)–图9(c)所示,可以发现改进后的算法无论在内河、湖泊以及海面背景下均能精准地检测出大量小目标,没有明显的误检、漏检情况,展现出较好的鲁棒性和泛化性,能够很好地满足实际场景下的检测需求。

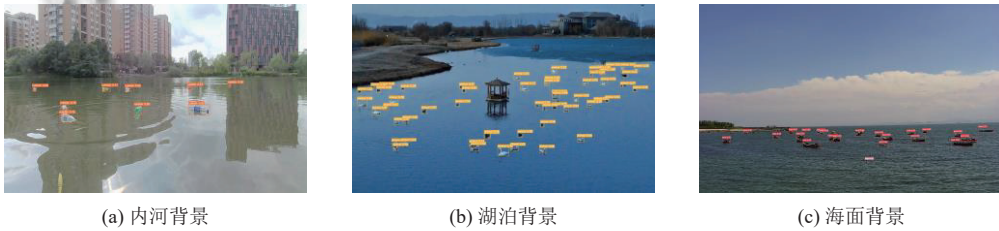


图9 不同拍摄背景检测效果

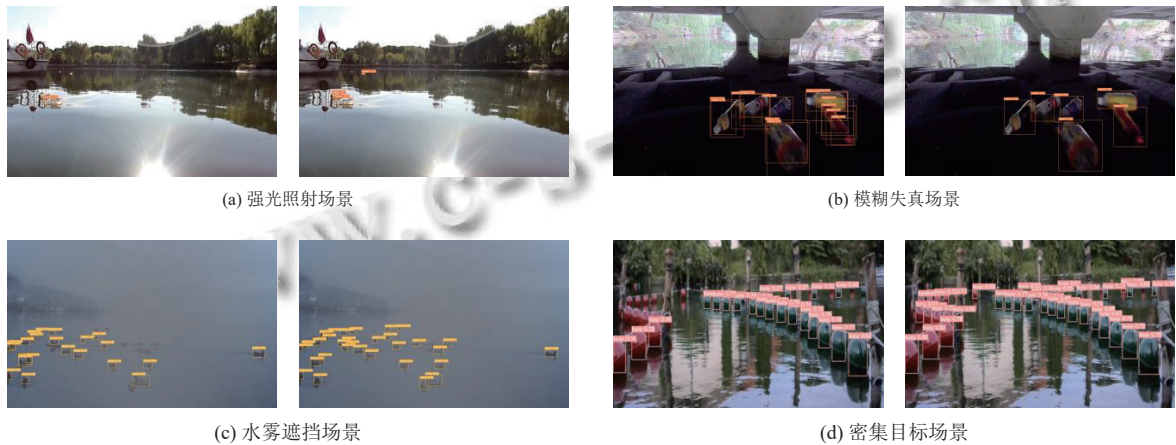


图10 检测效果对比

4 结论与展望

本文提出了一种改进 YOLOv8 的水面小目标检测算法,减轻了水面目标检测中的噪声干扰,避免了对水面小目标的误检和漏检的问题。通过引入 BiFormer 双层路由注意力机制构建 C2fBF 模块、添加小目标检测头,

为了进一步评估本文算法的检测性能,本文还选取了强光照射、模糊失真、水雾遮挡以及存在密集目标场景下的图像分别对基线模型 YOLOv8n 和本文提出的 YOLOv8-WSSOD 算法进行对比测试,如图10所示,每个子图中,左图为 YOLOv8n 算法的检测效果,右图为 YOLOv8-WSSOD 算法的检测效果。图10(a)是在强光照射场景下的检测对比,图10(b)是在光线昏暗条件下模糊失真图像的检测对比,可以看出改进后的算法受光线变化的影响很小,对环境噪声具备很好的抗干扰能力,与原始 YOLOv8n 算法相比,能精确地检测出水面目标的类别和位置。除此之外,改进后的算法也能精准地检测到水面上受遮挡或者重叠的小目标。图10(c)是在水雾遮挡场景下的检测对比,图10(d)是对有大量水面浮标重叠图像的检测对比,可以发现 YOLOv8-WSSOD 算法相比 YOLOv8n,学习到了更充分更细节的小目标特征,水雾遮挡或者目标重叠并没有影响模型对水面目标的判断与检测,极大地缓解了漏检问题。

替换 MPDIoU 损失函数等一系列改进方法,本文提出的 YOLOv8-WSSOD 算法比原始 YOLOv8n 算法分别提升了 4.6% 的 $mAP@0.5$ 和 2.2% 的 $mAP@0.5:0.95$,并保证了检测速度,具有较强实用性。然而本文所提出算法的计算量与原 YOLOv8n 模型相比并没有很好地降

低,在减少模型的参数量和计算负担上还有优化空间,故下一步还需对网络结构进一步改进,使其更加轻量高效。

参考文献

- 1 Zhang LL, Zhang Y, Zhang Z, *et al.* Real-time water surface object detection based on improved Faster R-CNN. *Sensors*, 2019, 19(16): 3523. [doi: [10.3390/s19163523](https://doi.org/10.3390/s19163523)]
- 2 刘伟,王源楠,江山,等.基于Mask R-CNN的水面漂浮物识别方法研究. *人民长江*, 2021, 52(11): 226–233. [doi: [10.16232/j.cnki.1001-4179.2021.11.037](https://doi.org/10.16232/j.cnki.1001-4179.2021.11.037)]
- 3 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. *广西大学学报(自然科学版)*, 2021, 46(6): 1569–1578. [doi: [10.13624/j.cnki.issn.1001-7445.2021.1569](https://doi.org/10.13624/j.cnki.issn.1001-7445.2021.1569)]
- 4 王林,汪钰婷.基于加强特征融合的轻量化船舶目标检测. *计算机系统应用*, 2023, 32(2): 288–294. [doi: [10.15888/j.cnki.csa.008948](https://doi.org/10.15888/j.cnki.csa.008948)]
- 5 Deng LW, Liu Z, Wang JD, *et al.* ATT-YOLOv5-Ghost: Water surface object detection in complex scenes. *Journal of Real-time Image Processing*, 2023, 20(5): 97. [doi: [10.1007/s11554-023-01354-z](https://doi.org/10.1007/s11554-023-01354-z)]
- 6 陈任飞,彭勇,吴剑,等.基于深度学习的水面漂浮物智能检测方法. *工程科学与技术*, 2023, 55(3): 165–174. [doi: [10.15961/j.jsuese.202101158](https://doi.org/10.15961/j.jsuese.202101158)]
- 7 张晓鹏,许志远,曲胜,等.基于改进YOLOv5深度学习的海上船舶识别算法. *大连海洋大学学报*, 2022, 37(5): 866–872. [doi: [10.16535/j.cnki.dlhyxb.2022-104](https://doi.org/10.16535/j.cnki.dlhyxb.2022-104)]
- 8 Zhu L, Wang XJ, Ke ZH, *et al.* BiFormer: Vision Transformer with bi-level routing attention. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 10323–10333. [doi: [10.1109/cvpr52729.2023.00995](https://doi.org/10.1109/cvpr52729.2023.00995)]
- 9 Li HL, Li J, Wei HB, *et al.* Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv:2206.02424*, 2022.
- 10 Ma SL, Xu Y. MPDIoU: A loss for efficient and accurate bounding box regression. *arXiv:2307.07662*, 2023.
- 11 Feng CJ, Zhong YJ, Gao Y, *et al.* TOOD: Task-aligned one-stage object detection. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE, 2021. 3490–3499. [doi: [10.1109/iccv48922.2021.00349](https://doi.org/10.1109/iccv48922.2021.00349)]
- 12 Li X, Wang WH, Wu LJ, *et al.* Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 1763.
- 13 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 14 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16×16 words: Transformers for image recognition at scale. *Proceedings of the 9th International Conference on Learning Representations*. ICLR, 2020.
- 15 Carion N, Massa F, Synnaeve G, *et al.* End-to-end object detection with Transformers. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020. 213–229. [doi: [10.1007/978-3-030-58452-8_13](https://doi.org/10.1007/978-3-030-58452-8_13)]
- 16 Cheng YW, Zhu JN, Jiang MX, *et al.* FloW: A dataset and benchmark for floating waste detection in inland waters. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 10933–10942. [doi: [10.1109/iccv48922.2021.01077](https://doi.org/10.1109/iccv48922.2021.01077)]
- 17 Zhou ZG, Sun JE, Yu JB, *et al.* An image-based benchmark dataset and a novel object detector for water surface object detection. *Frontiers in Neurorobotics*, 2021, 15: 723336. [doi: [10.3389/fnbot.2021.723336](https://doi.org/10.3389/fnbot.2021.723336)]
- 18 Zheng ZH, Wang P, Liu W, *et al.* Distance-IoU loss: Faster and better learning for bounding box regression. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020. 12993–13000. [doi: [10.1609/aaai.v34i07.6999](https://doi.org/10.1609/aaai.v34i07.6999)]
- 19 Rezaatfighi H, Tsoi N, Gwak JY, *et al.* Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 658–666. [doi: [10.1109/cvpr.2019.00075](https://doi.org/10.1109/cvpr.2019.00075)]
- 20 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015. 91–99.
- 21 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 21–37.
- 22 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 7464–7475. [doi: [10.1109/cvpr52729.2023.00721](https://doi.org/10.1109/cvpr52729.2023.00721)]
- 23 Zhu XK, Lyu SC, Wang X, *et al.* TPH-YOLOv5: Improved YOLOv5 based on Transformer prediction head for object detection on drone-captured scenarios. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops*. Montreal: IEEE, 2021. 2778–2788. [doi: [10.1109/iccvw54120.2021.00312](https://doi.org/10.1109/iccvw54120.2021.00312)]

(校对责编:孙君艳)