

基于改进的 YOLOv7 河道漂浮物小目标检测模型

王国明, 宋 健

安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001

摘要:目的 针对机器人清理河道漂浮物作业中的计算能力有限和河道环境复杂造成的小目标漏检等问题, 在 YOLOv7 基础上提出了一种兼顾精度与轻量化的 YOLOv7 目标检测模型。方法 该模型是在原 YOLOv7 上增加一个大小为 160×160 的小目标检测层, 提升对小目标的特征学习能力, 并在原网络模型的颈部引入 BiFormer 注意力机制, 增强对小目标的检测性能, 以减少在河道漂浮物检测过程中的漏检现象; 其次用 PELAN 替换主干网络的 ELAN 模块, 减少算法的参数数量和运算量, 提高网络检测速度; 在此基础上, 引入 MPDIoU 损失函数, 提升边界框回归的收敛速度和精度, 提高网络模型的鲁棒性。结果 该改进模型在欧卡智船的 FloW 子数据集下的 mAP 达到 71.5%, 相比原 YOLOv7 网络模型提升了 4.7%, 模型参数量和运算量分别降低 11% 和 7.3%。结论 通过对比实验表明: 整体效果优于原网络模型与传统经典目标检测网络模型, 在提高精度的同时, 能够减少模型复杂度, 更贴合于实际应用。

关键词:小目标检测; YOLOv7 网络模型; BiFormer; 轻量化; 损失函数

中图分类号:X52; TP391.41 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2026.0003.008

A Small Target Detection Model for River Floating Objects Based on Improved YOLOv7

WANG Guoming, SONG Jian

School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, Anhui, China

Abstract: Objective In the operation of cleaning river floating objects using robots, limited computational capacity and complex river environments often lead to the missed detection of small targets. To address these issues, an improved YOLOv7 object detection model that balances accuracy and lightweight design is proposed based on YOLOv7. **Methods** First, a small target detection layer with a size of 160×160 was added to the original YOLOv7 to enhance the feature learning ability for small objects. The BiFormer attention mechanism was introduced into the neck of the original network model to improve detection performance for small targets, thereby reducing missed detections during river floating object detection. Second, the ELAN module in the backbone was replaced with PELAN to reduce the number of parameters and computational overhead, thus increasing the detection speed. On this basis, the MPDIoU loss function was incorporated to accelerate the convergence and improve the accuracy of bounding box regression and enhance the robustness of the network. **Results** The improved model achieved a mean average precision (mAP) of 71.5% on the FloW subset from Oka-Zhibo, which was 4.7% higher than that of the original YOLOv7 model. Moreover, the number of parameters and computational cost were reduced by 11% and 7.3%, respectively. **Conclusion** Comparative experiments show that the overall performance of the improved model is better than that of the original network model and traditional classic target-detection network models. The proposed model can reduce the model complexity while improving the

收稿日期:2024-04-12 修回日期:2024-08-18 文章编号:1672-058X(2026)03-0061-09

基金项目:国家自然科学基金(青年)基金(62102003)项目资助;安徽省大学生创新创业基金(S202310361157, S202310361161)项目资助。

作者简介:王国明(1965—),男,安徽阜阳人,副教授,硕士研究生导师,从事图形图像处理研究。

通信作者:宋健(1998—),男,硕士研究生,从事目标检测研究。Email:3389238872@qq.com.

引用格式:王国明,宋健.基于改进的 YOLOv7 河道漂浮物小目标检测模型[J].重庆工商大学学报(自然科学版),2026,43(3):61-69.

Wang Guoming, Song Jian. A small target detection model for river floating objects based on improved YOLOv7[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(3): 61-69.

accuracy, making it more suitable for practical applications.

Keywords: small target detection; YOLOv7 network model; BiFormer; lightweight; loss function

近年来,随着深度学习的不断发展以及计算机软硬件性能的提升,基于深度学习的目标检测技术在工业场景中已经得到了进一步应用。传统目标检测算法主要由手工进行特征构建,导致很多任务分多次解决,并且存在速度慢、精度低等问题。随着卷积神经网络的兴起,图像数据集的大量出现和图像处理速度的大幅提升,深度学习目标检测技术得到了前所未有的突破,能够在某些特定的应用领域代替人工或其他传感检测,甚至优于人眼识别,成为目前目标检测^[1]方向的研究主流。

在无人船技术发展过程中,将目标检测技术应用于河道漂浮垃圾的检测一直是研究者关注的问题之一,其主要挑战在于如何有效地检测出图像中较小的目标。由于小目标^[2-3]尺寸较小,其特征信息很难被准确检测和识别出来,因此,小目标检测成为当前具有挑战性的研究方向。

目前目标检测技术主要分为两类:单阶段检测(如 YOLO^[4-6]、SSD^[7])和双阶段检测(如 Faster R-CNN^[8])。单阶段检测通过一次网络前向传播同时完成检测和分类,以快速响应著称,但可能在精度上有所下降。而双阶段检测则首先定位候选区域,然后对这些区域进行详细分类和定位来提高检测精度,但这以牺牲速度为代价。且在小目标检测领域中,一部分主要集中在加大模型,这使得模型参数量增多,进而导致模型检测速度降低;另一部分主要集中在轻量化模型,这限制它们提取深层特征的能力,导致检测精度下降,尤其是在小目标或者复杂场景中。Zhang 等^[9]使用 MobileNetv3 架构替换 YOLOv4 的骨干网络来实现目标检测,虽然使得模型更为轻量化,但其感受野较小,在处理中大型物体时表现不佳。戚玲琰等^[10]通过改进 MPCConv 模块并引入 ACmix 注意力机制,能够减少特征损失并提升检测性能,但在替换损失函数时,模型收敛速度较慢。Deng 等^[11]基于 Mask R-CNN,通过添加空洞卷积引入特征金字塔网络,提升了目标检测精度,但是模型参数量较多,无法进行实时检测。Zou 等^[12]基于 YOLOv5 和 Ghost 轻量化模块,引入 SE 注意力机制,虽有较好的模型性能,但是检测速度还存在一定不足。郝琨等^[13]将 CBAM^[14]机制集成到不同尺寸的检测层中,提出了 CPM-YOLOv3 模型,这一改进有效减少了模型的参数量;然而,该模型在小目标特征提取方面并未实现显著提升,这导致其在水下小目标检测的精度上仍有提升空间。

上述研究成果对小目标在计算机视觉领域提供了

一些重要的解决思路,但仍然存在一些需要改进的地方。一些算法虽然在小目标检测方面取得了进展,但它们忽视了实际应用中物体尺寸的多样性,未能充分考虑不同物体的检测需求。此外,尽管一些研究已经对模型参数进行了优化,但检测准确性仍有提升空间。为了解决上述问题,本文提出一种新的 YOLOv7^[15]目标检测模型。通过增加目标检测层,使其在浅层网络获得更多目标的位置信息,提高对小目标的检测率。针对小目标检测的误检情况,引入 Transformer 下的 BiFormer 注意力机制,用更少的计算量捕获更多的特征,保留更多的细节信息,有效提高了模型对小目标的检测精度。PConv 卷积^[16]技术可以大幅度降低检测网络的重量,从而极大地改善其性能。它不仅可以降低冗余计算,而且还可以节省大部分的资源,从而更加有效地提取空间特征,并且拥有良好的计算效率以及较低的模型参数。最后通过改进损失函数,使用基于最小点距离的新型边界框相似度比较度量,简化了两个边界框之间的相似性比较,具有更高的边界框回归效率和精度,能够更好地提升模型的鲁棒性。

1 检测系统设计及其改进

1.1 YOLOv7 相关技术

YOLOv7 是一种先进的目标识别算法,它的架构包括一个输入模块,一个 Backbone,一个 Neck,一个 Head,它们分别负责处理各种信息,并且能够根据不同的需求调整参数。该算法的架构如图 1 所示。

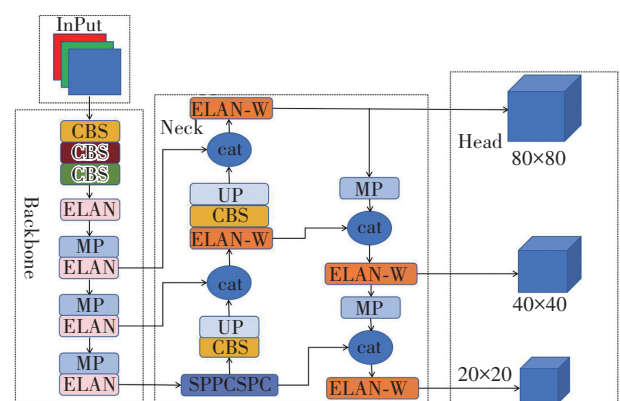


图 1 YOLOv7 网络结构

Fig. 1 YOLOv7 network structure

Backbone 网络的核心是 CBS、ELAN 和 MPCConv。在这些结构之间,CBS 由卷积、BN(批量归一化)和 SiLu 激活函数构成,主要用于提取特征。ELAN 卷积层通过多个 CBS 连接,组成一个高效的远程注意网络,能

够有效处理信息的复杂性特征。MP 卷积主要由 CBS 和 MaxPool 组成,分别提取相关区域的所有信息和最大信息。最后,通过使用 Concat 操作将两路特征进行融合,从而有效增强特征提取功能。Backbone 各个模块如图 2 所示。

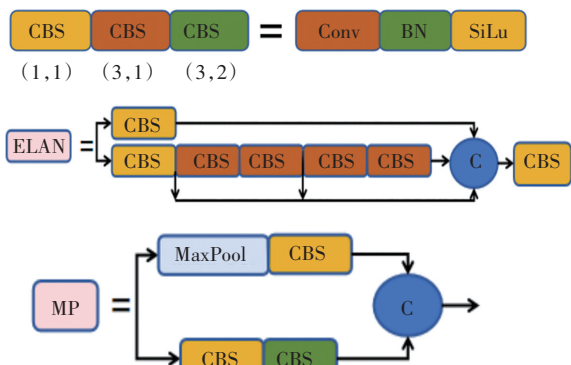


图 2 Backbone 模块结构图

Fig. 2 Backbone module structure diagram

通过骨干网络和颈部网络的特征融合,图像可以获得更加丰富的信息。其中,头部网络采用 SPPCSPC (Spatial Pyramid Pooling Cross Stage Partial Csp) 金字塔结构,可以扩大感受野,从而使其能够处理多尺寸输入;而颈部网络则采用 PAFPN (Path Aggregation Feature

Pyramid Network) 结构^[17],可以从底层信息开始,沿着自下而上的路径传递到深层网络,从而实现更加丰富的信息处理功能。通过 ELANW,可以将多个层次的特征融合在一起,并且具有不同的采样倍率,从而实现与 ELAN 卷积层相似的功能。Neck 各个模块结构如图 3 所示。头部网络负责对 anchor 进行分类和回归,输出最终的预测框。

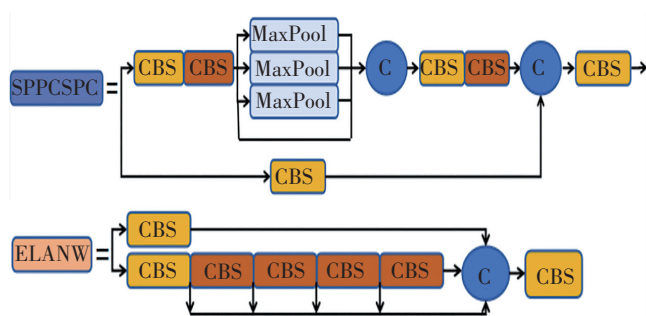


图 3 Neck 模块结构图

Fig. 3 Neck module structure diagram

1.2 YOLOv7 目标检测模型的改进

为了提升 YOLOv7 在河道漂浮物小目标检测中的准确性并减少漏检和误检情况,模型提出几项关键改进,改进后的 YOLOv7 网络结构如图 4 所示。

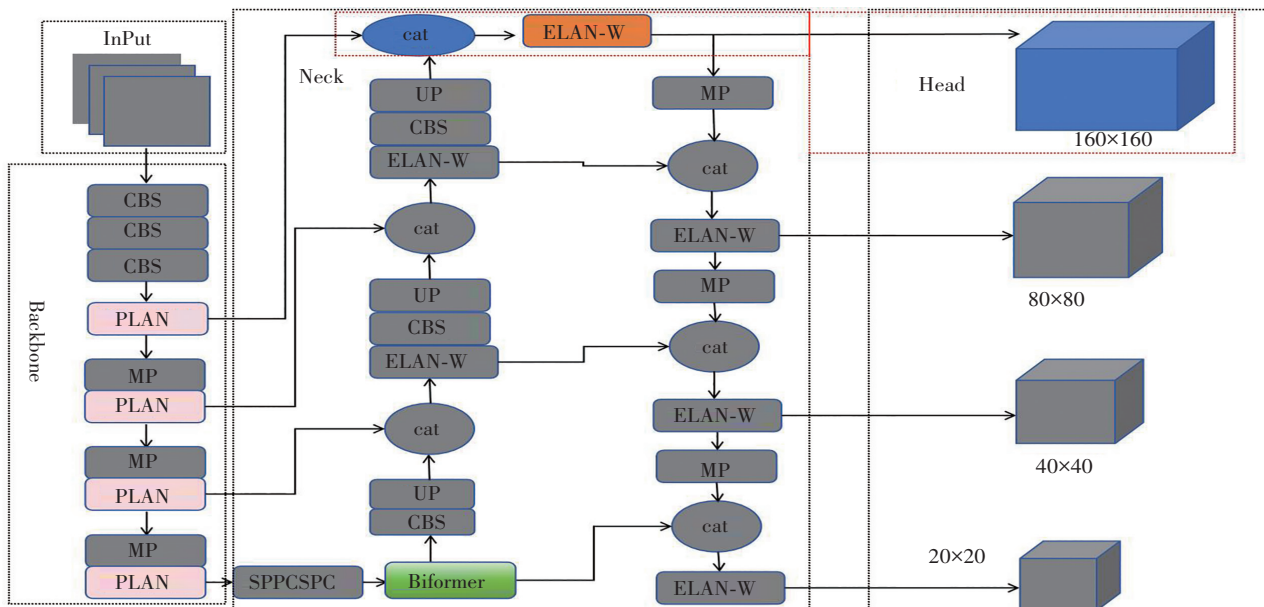


图 4 改进的 YOLOv7 网络结构

Fig. 4 Improved YOLOv7 network structure

1.2.1 小目标检测层

为了应对河道场景中目标结构和尺寸的多样性,以及提高 YOLOv7 在小目标检测方面的性能,本研究对模型进行了创新性的改进。具体来说,通过增强 Backbone 网络的特征提取能力和 Neck 网络的特征融合策略,YOLOv7 能够生成 3 个不同尺度的特征图,以适应复杂的河道目标检测任务。

针对河道中目标的多变性,尤其是小目标检测的挑战,研究中引入了一个新的小目标检测层。这个新增层的输出尺寸为 160x160 像素,专门为提升 YOLOv7 对小目标的检测精度而设计。

新增的小目标检测层位于现有特征融合模块之后,由两大部分构成:一部分是直接来自 Backbone 网络首个 ELAN 模块的特征图,另一部分则是通过 Neck 网

络进行上采样至 160×160 像素的特征图。这两部分特征图经过 ELAN-W 模块的进一步融合处理,形成了一个综合高维度信息的复合特征图。

这个复合特征图虽然具有较小的感受野,但它提供了丰富的位置信息,这对于精确定位小目标至关重要。因此,通过这种结构上的优化,改进后的 YOLOv7 模型在处理河道漂浮物小目标检测问题时,能够显著降低误检和漏检的发生率,从而提升检测的准确性和鲁棒性。

在图 4 中,新增的小目标检测层以红色虚线框标识,清晰展示了其在改进后 YOLOv7 网络结构中的位置。通过这种结构优化,改进后的 YOLOv7 模型能够更有效地处理河道漂浮物小目标的检测问题,显著提升检测的准确性和鲁棒性。

1.2.2 BiFormer 注意力机制

BiFormer 作为 Transformer 模型的一个创新变体,引入了一种先进的动态注意力机制,通过双层路由策略实现了更为灵活和高效的内容感知计算分配,赋予了模型动态查询感知的稀疏性特质。BiFormer 结构如图 5 所示。

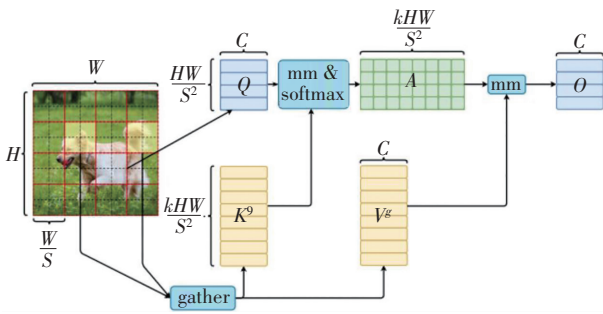


图 5 BiFormer 模块

Fig. 5 BiFormer module

BiFormer 的工作流程大致可分为 3 个关键步骤:

(1) 特征图的区域划分与键值对生成。首先,模型接收一个特征图作为输入,并将该图划分为若干个区域;随后,通过线性映射操作,为每个区域生成对应的查询(Query)、键(Key)和值(Value)张量,为后续的注意力计算奠定基础。

(2) 有向图的构建与键值对关联。接着,利用邻接矩阵构造出一个有向图,该图能够捕捉不同区域之间的参与关系。在此基础上,模型识别出相互关联的键值对,为实现精确的注意力分配提供了结构化的信息。

(3) 细粒度的路由索引与注意力机制应用。最后,模型使用路由索引矩阵指导从一个区域到另一个区域的信息流动。这一过程涉及细粒度的令牌到令牌注意力机制。通过这种方式,BiFormer 能够更加精确地筛选出与目标相关的特征,同时忽略无关信息,从而提

高模型的计算效率和检测性能。

BiFormer 的动态注意力机制是其核心优势之一,它通过动态调整注意力分配,有效避免了对无关信息的无效搜索,同时保留了对目标特征的敏感度,因此将 BiFormer 与 YOLOv7 联合应用进行优化具有可行性。

模型首先生成 Q, K, V 键值对,这些键值对是构建注意力机制的基础。通过计算这些键值对,模型能够在区域级别上识别出最优的路由路径,这一过程涉及对特征图的粗略分析,以确定哪些区域在当前上下文中最为重要。

筛选出的最优路由随后与其他路径相结合,形成一个更为全面的路由网络。在此基础上,模型进一步细化其注意力机制,通过令牌到令牌的关注策略($K^s V^s$),对每个区域内的元素进行更精细的分析和考量。

$$K^s = \text{gather}(K, I) \quad (1)$$

$$V^s = \text{gather}(V, I) \quad (2)$$

在这个过程中,模型收集并构建了 K^s, V^s 键值张量,该张量记录了每个区域与其他区域之间的相关性。具体来说,对于键值张量的第 i 行,它包含了第 i 个区域与前 K 个最相关区域的索引映射。这里的 K 是一个预设参数,表示在考虑的前 K 个最相关的邻域窗口中,每个区域都会收集与之相关联的键值对。

通过上述步骤,BiFormer 模型不仅能够识别出图像中的关键区域,还能够在这些区域内实现精确的注意力分配,从而为小目标检测任务提供一种高效的特征提取和信息整合机制。这种方法特别适用于处理那些在空间上相互遮挡或密集排列的小目标,能够显著提高检测的准确性和鲁棒性。

1.2.3 骨干网络轻量化

为了实现检测网络的轻量化,并确保快速而准确的检测性能,本研究采纳了 FasterNet 架构中提出的 PConv 模块。该模块被用于替代主干网络中原有的 ELAN 模块内 3×3 大小、步长为 2 的标准卷积层。通过这种替换,PConv 模块在减少参数数量和浮点运算次数(FLOPs)方面发挥了关键作用,从而降低了模型的复杂度,加快了检测速度,同时保持了检测结果的高准确率。尽管 GhostNet^[18-19] 和其他轻量级网络能够通过有效减小参数和浮点计算,来取代传统卷积神经网络,然而,它们频繁的内存访问和深层卷积操作,使得相关通道数量翻了一番,进一步提升了系统的性能,但也带来了一定的延迟,使得整个系统的计算效率急剧下滑,最终造成严重的检测精度损失。PConv 卷积可以大幅降低冗余计算量,并且可以节省大量的内存,从而大大增强了空间特征的捕捉能力,克服了传统卷积及其他深度学习技术的缺陷。它的每秒浮点操作次数(FLOPS)

也可以得到显著改善,从而使得整个系统的性能大大增强。因此,本文基于 PConv 算子提出轻量化 PELAN。PELAN 结构如图 6 所示,PConv 的模块结构如图 7 所示。

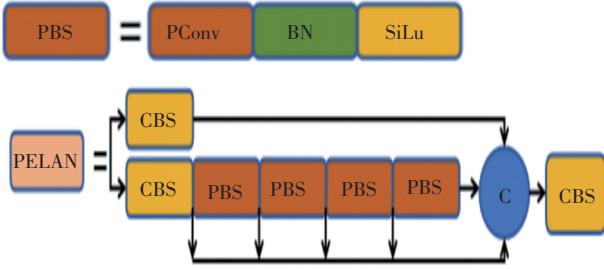


图 6 PELAN 结构

Fig. 6 PELAN structure

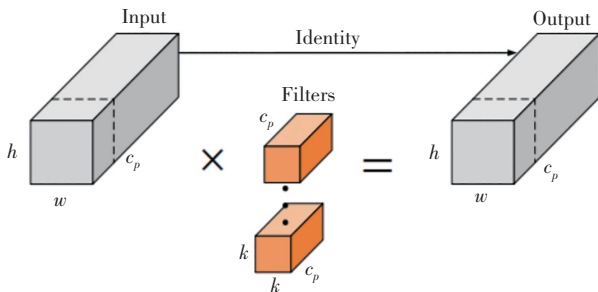


图 7 PConv 模块结构图

Fig. 7 Structure diagram of the PConv module

为了实现高效的内存访问,模型采用了 PConv 模块(图 7),将一部分输入通道直接整合为特征图的一部分。在这种设计下,PConv 的浮点运算次数(FLOPs)显著降低。具体来说,如果将 PConv 模块中的卷积通道数设置为输入通道数 c 的 $1/4$,其 FLOPs 将降至常规卷积(Conv)的 $1/16$,内存访问量也相应减少至常规卷积的 $1/4$ 。这样的设计使得 PConv 在保持低 FLOPs 的同时,相比深度可分离卷积(DWConv)或分组卷积(GConv)展现出更高的性能。

此外,PConv 模块能够更充分地利用计算硬件的能力,在提取空间特征的同时保持了良好的效果。因此,通过应用 PConv 模块,不仅可以维持网络的轻量化,还能增强特征图的生成质量,这对于小目标检测模型的训练和推理过程来说是一个显著的优势。

1.2.4 损失函数

损失函数是评价模型预测准确性的重要工具,它不仅反映了预测结果与实际标注之间的偏差,而且对网络的快速收敛起到了至关重要的作用。通过引导网络在训练过程中最小化这一差异,损失函数显著提升了模型的预测精度。

在原始的 YOLOv7 网络模型中,采用 CIoU (Complete Intersection over Union) 损失函数作为训练目标。CIoU 损失函数综合考虑了边界框的中心点距离、长宽比以及重叠区域的比例,这使得它能够有效地处理目标检

测中的定位问题。然而,CIoU 损失函数在处理低质量锚框时收敛困难,这限制了它在某些情况下的性能表现。其计算公式如下:

$$L_{CIoU} = 1 - L_{IoU} + \frac{\rho^2(b, b_{gt})}{c^2} + av \quad (3)$$

$$a = \frac{v}{1 - L_{IoU} + v} \quad (4)$$

$$v = \frac{4}{\pi^2} \left[\arctan\left(\frac{w_{gt}}{h_{gt}}\right) - \arctan\left(\frac{w}{h}\right) \right]^2 \quad (5)$$

$$\frac{\partial v}{\partial w} = \frac{8}{\pi^2} \left(\tan^{-1} \frac{w}{h} - \tan^{-1} \frac{w_{gt}}{h_{gt}} \right)^2 \frac{h}{h^2 + w^2} \quad (6)$$

$$\frac{\partial v}{\partial h} = \frac{8}{\pi^2} \left(\tan^{-1} \frac{w}{h} - \tan^{-1} \frac{w_{gt}}{h_{gt}} \right)^2 \frac{w}{h^2 + w^2} \quad (7)$$

其中,符号 b 代表预测框,即模型预测出的物体边界框;符号 b_{gt} 代表真实框,即标注的正确物体边界框;符号 c 表示预测框和真实框共同构成的最小包围区域的对角线长度;符号 a 是一个平衡参数,用于调整损失函数中不同因素的权重;符号 v 用于表示预测框与真实框的长宽比是否一致。根据式(3),当 v 取值为 0 时,意味着预测框与真实框的长宽比是相等的,此时长宽比的一致性不会对损失函数的计算产生影响。这表明在某些情况下,损失函数可能会忽略长宽比的匹配,这可能会影响模型在处理不同形状物体时的性能。因此,这里的长宽比并没有起到作用,CIoU 损失函数将无法被准确地描述,进而阻碍模型收敛。且在式(6)和式(7)中, w 、 h 和 v 是一对相反数,并不会同时增大或减小。

因此,本文引入 MPDIoU 损失函数,使用基于最小点距离的新型边界框相似度比较度量。相较于原网络中的 CIoU,其具有更高的边界框回归效率和精度,通过最小化预测边界框和真实边界框之间的左上和右下点距离,MPDIoU 损失函数所用的参数如图 8 所示。

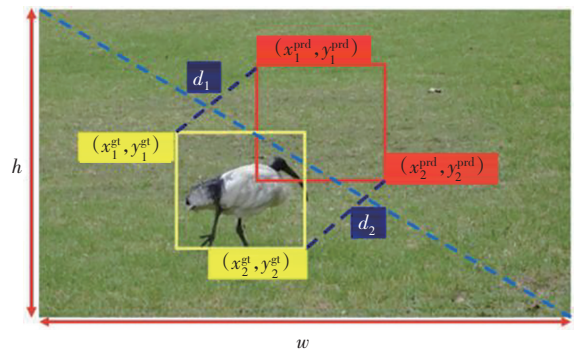


图 8 MPDIoU 损失函数

Fig. 8 MPDIoU loss function

MPDIoU 的主要计算公式如下:

$$d_1^2 = (x_1^{prd} - x_1^{gt})^2 + (y_1^{prd} - y_1^{gt})^2 \quad (8)$$

$$d_2^2 = (x_2^{\text{prd}} - x_2^{\text{gt}})^2 + (y_2^{\text{prd}} - y_2^{\text{gt}})^2 \quad (9)$$

$$f_{\text{MPDIoU}} = L_{\text{IoU}} \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (10)$$

$$L_{\text{MPDIoU}} = 1 - f_{\text{MPDIoU}} \quad (11)$$

其中,边界框回归损失函数的所有参数都可以通过 4 个点的坐标来计算。 $(x^{\text{prd}}, y^{\text{prd}})$ 表示预测框的左上和右下顶点坐标, $(x^{\text{gt}}, y^{\text{gt}})$ 表示真实框的左上和右下顶点坐标, d 表示真实框与预测框对应顶点的距离, w 、 h 分别表示图片的宽度与高度。MPDIoU 简化了两个边界框之间的相似性比较,相较于 CIoU,更适用于重叠或非重叠的边界框回归。并且,如果预测边界框和真实边界框的宽高比相同,则在真实边界框内的预测边界框 L_{MPDIoU} 值低于在真实边界框外的预测框。这一特性保证了边界框回归的准确性,使得预测的边界框具有较少的冗余性。

2 实验设计与结果分析

2.1 实验环境与参数设置

实验所用的网络环境配置包括 16 个虚拟 CPU 核心,搭载 Intel(R) Xeon(R) Platinum 8350C 处理器,主频为 2.60 GHz,配备 42 GB 的内存。网络模型的训练过程是在一块 NVIDIA GeForce RTX 3090 显卡上完成的。系统运行的是 Windows 11 操作系统,使用 Python 3.8 作为编程语言,依托于 PyTorch 1.9.0 作为深度学习框架,并利用 CUDA 11.7 提供加速计算支持。这样的配置为复杂网络模型的训练和深度学习任务提供了强大的计算能力和高效的运行环境。

本文以 YOLOv7 为基础模型,训练数据量为 300 个 epoch,初始学习率设置为 0.01,批次大小设置为 16,动量参数设置为 0.937。

2.2 数据集

在本次实验中,所采用的数据集为 FLoW 数据集,这是由欧卡智船提供的,专门用于模拟无人船视角下河流漂浮垃圾检测的专用数据集。FLoW 数据集的独特之处在于,它是全球首个针对真实内河水域场景,综合多传感器和多种天气条件下的内河无人船操作而设计的。特别值得一提的是,FLoW 的子数据集 FLoW-Img 小目标占比超过一半,涵盖了各种环境、地点和季节等信息。该研究使用 2 000 张高质量的图像,每张图像的像素尺寸均在 720×720 之间,按照 7:1:2 的比例对这些数据进行分组,并对其进行验证和测试。

2.3 评价指标

为了全面评估模型性能并准确衡量实验结果的改进,采用多种指标进行综合考量。参数量是模型参数总数,影响模型复杂度;计算量 (FLOPs) 是模型前向传播的浮点运算次数,影响计算效率;准确率 ($\zeta_{\text{precision}}$) 是正确正样本占有所有正确预测的比例,其中 N_{TP} 是正确检

测的正样本数量, N_{FP} 是错误检测为正的负样本数量;召回率 (ζ_{Recall}) 是正确正样本占有所有实际正样本的比例,其中 N_{FN} 是漏检的负样本数量;平均精度 (ζ_{mAP}) 是综合考虑 $\zeta_{\text{precision}}$ 和 ζ_{Recall} ,衡量模型在不同类别上的平均性能, ζ_{AP} 是相应类别的精度, N 是检测类别的总数;FPS 为每秒检测多少帧图片。相应的计算公式如下:

$$\zeta_{\text{Precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (12)$$

$$\zeta_{\text{Recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (13)$$

$$\zeta_{\text{mAP}} = \frac{\zeta_{\text{AP}}}{N} \quad (14)$$

2.4 融合注意力机制比较

为了进一步验证注意力机制改进的有效性,对添加的注意力机制进行更换,包括 SE、ECA、CA、CBAM 和 Biformer。对比分析不同注意力机制给模型检测效果带来的影响,训练过程中采用相同参数,并且在 FLoW 数据集上进行实验。结果如表 1 所示。

表 1 注意力机制消融实验结果

Table 1 Ablation study results of attention mechanisms

Model	Params/MB	FLOPs/G	mAP/%	FPS
—	37.2	105.3	66.8	71.7
SE	38.1	105	63.9	68.4
ECA	37.6	104.7	66.4	69.8
CA	38.2	105.9	67.1	70.1
CBAM	37	105.5	68.1	71.3
Biformer	37.5	99.3	68.9	71.5

SE (Squeeze-and-Excitation) 是一种经典的通道注意力机制,它通过压缩和激励过程来强化特征通道的重要性。ECA (Efficient Channel Attention) 作为通道注意力的另一种形式,通过有效捕捉跨通道的关联性来增强特征表示,但忽略了空间位置信息。CA (Convolutional Attention) 将位置信息整合到通道注意力中,通过两个特征编码步骤分别处理不同空间方向的特征,从而生成融合通道和空间信息的权重。CBAM (Convolutional Block Attention Module) 综合通道和空间注意力机制的优势,独立地对通道和空间权重进行建模,既强化了通道间的关系,又考虑了空间上的相互作用,实现了对特征的全面优化。

实验结果揭示了不同注意力机制对模型性能的具体影响。引入 SE 和 ECA 注意力机制的模型在检测准确率上分别遭遇了 2.9% 和 0.4% 的下降,并且模型参数量有所增加。这表明在当前数据集上,这两种机制并未能有效地提升性能。相比之下,当模型融合 CA、

CBAM 和 Biformer 注意力机制时,检测准确率分别获得了 0.3%、1.3%和 2.1%的提升,对于后两个注意力机制,检测精度明显提升。而 Biformer 机制则在提高精度的同时,还对模型计算进行了优化,使得网络能够更精准地捕捉目标区域的细节特征,更好地适应河道垃圾检测的实际需求。综合来看,Biformer 注意力机制的引入不仅加快了检测速度,而且有效提升了模型的检测精度,这使其在实际应用中更具优势。

2.5 消融实验

消融实验旨在验证所提出的改进措施对 YOLOv7 模型性能提升的有效性。实验共设置 8 组,每组均在保持输入图像、训练参数等条件一致的前提下,通过引入不同的改进措施进行测试。这些改进措施包括小目标检测层(Head)、Biformer 注意力机制、MPDIoU 损失函数以及 PELAN 模块。实验结果汇总在表 2 中,其中“√”表示该实验组采纳了相应的改进措施。

表 2 消融实验结果
Table 2 Ablation study results

组 别	Head	Biformer	MPDIoU	PELAN	FLOPs/G	Params/MB	mAP%	FPS
1					105.3	37.2	66.8	71.7
2	√				119.6	37.7	68.9	67.1
3		√			99.3	37.5	70.0	71.5
4			√		105.3	37.2	67.3	72.9
5				√	84.9	32.7	65.9	83.4
6	√	√			114.6	37.7	71.8	67
7	√	√	√		114.6	37.7	72.2	69
8	√	√	√	√	97.6	33.2	71.5	77.2

表 2 的消融实验结果展示了 YOLOv7 算法经过不同改进后的性能提升情况。原始 YOLOv7,mAP 为 66.8%;加入小目标检测层,mAP 提升至 68.9%,扩大了小目标检测的范围;引入 Biformer 注意力机制,mAP 升至 70.0%,计算量减少,提高了检测精度和速度;采用 MPDIoU 损失函数,增强了模型的回归效率和鲁棒性,提升了检测精度;替换为 PELAN 模块,参数量减少,模型轻量化,检测速度提高,但 mAP 略有下降;结合小目标层与 Biformer,mAP 显著提高至 71.8%,证实 Biformer 的有效性;引入 MPDIoU 损失函数,mAP 达到 72.2%,展示了 MPDIoU 在提升精度上的贡献;最终改进模型,mAP 为 71.5%,比原始模型提高 4.7%,参数量和计算量分别减

少 11%和 7.3%,检测速度提升了 5.5,实现了精度提升和模型简化。这些改进显著提高了 YOLOv7 在小目标检测任务中的表现,同时优化了模型的计算效率。

这些实验结果表明,通过逐步集成小目标检测层、Biformer 注意力机制、MPDIoU 损失函数和 PELAN 模块,可以显著提高 YOLOv7 在小目标检测任务中的性能,并优化模型的计算效率和参数规模。

2.6 对比实验分析

为了更好地证明本文提出算法的可行性,将其与目前比较常用的算法:SSD、Faster R-CNN、YOLOv5 和各个版本的 YOLO 与其他模型在欧卡智舶数据集上进行对比,实验结果如表 3、表 4 所示。

表 3 对比实验结果
Table 3 Comparative experimental results

网络模型	输入尺寸/像素	mAP@0.5	mAP@0.5 : 0.95	Params/MB	FPS
SSD(VGG)	720×720	0.581	0.253	26.3	42
Faster R-CNN	720×720	0.402	0.150	—	19
YOLOv5	720×720	0.565	0.245	36.2	69.2
YOLOv7	720×720	0.668	0.279	37.2	71.7
YOLOv7-tiny	720×720	0.567	0.227	5.8	100.9
YOLOv8m	720×720	0.667	0.282	25.9	81.8
YOLOv9c	720×720	0.673	0.285	25.3	83.1
ours	720×720	0.715	0.304	33.2	77.2

表 4 YOLO 系列对比实验结果
Table 4 Comparative experiment results of YOLO series

网络模型	输入尺寸/像素	mAP@ 0.5	mAP@ 0.5 : 0.95	Params/MB	FPS
YOLOv7	720×720	0.668	0.279	37.2	71.7
YOLOv8n	720×720	0.593	0.235	3.2	108.9
YOLOv8s	720×720	0.622	0.259	11.1	89.2
YOLOv8m	720×720	0.667	0.282	25.7	81.8
YOLOv8l	720×720	0.677	0.285	43.7	78.3
YOLOv8x	720×720	0.689	0.291	66.9	62
YOLOv9t	720×720	0.611	0.273	2	122
YOLOv9s	720×720	0.638	0.279	7.1	95.4
YOLOv9m	720×720	0.679	0.289	20	88.7
YOLOv9c	720×720	0.673	0.285	25	83.1
YOLOv9e	720×720	0.692	0.295	58.3	69.8
ours	720×720	0.715	0.304	33.2	77.2

改进后的算法在目标检测的准确性上取得了显著进步,具体表现在与其他主流算法的比较中。在 mAP@ 0.5 的评估指标上,该算法比 SSD 提升了 13.4%,相较于 Faster R-CNN 取得了更大的改进幅度,比 YOLOv5 提高了 15%,并且相对于 YOLOv7、YOLOv8x 和 YOLOv9e 分别提升了 4.7%、2.6% 和 2.3%;对比其他轻量化的版本,提升值将会更高,最高达到 14.8%。在 FPS 的指标上,该算法检测速度虽比 YOLO 其他轻量化版本略有不足,但能够满足实际应用需求。

从 YOLO 系列及其轻量化版本的对比结果来看,虽然轻量化模型在减少参数量方面表现更佳,但在检测精度上通常有所折中。然而,经过研究和改进后的 YOLOv7 模型,不仅在参数量上有所减少,更重要的是,

在检测精度和速度上实现了显著提升。该模型的 mAP 值达到了 0.715,这一结果不仅在数值上具有明显优势,而且在实际应用中,更好地满足了对河道漂浮物进行检测和精确度识别的要求。

为了直观评估改进后的网络模型在目标检测任务中的性能,本实验直接从实际场景拍摄中选取若干图像进行测试,对 YOLOv7 原始网络模型和改进模型进行两组对比测试。图 9 用来测试在密集图片中目标的检测效果,图 10 用来测试远距离小目标检测情况。在图 9 当中,原 YOLOv7 网络模型在检测中漏检两个漂浮瓶,并且错检了一个,改进后的网络模型则全部检测出;在图 10 中,远距离漂浮瓶特征少,更难提取有效特征,但改进后的效果比原 YOLOv7 网络模型多检测出一个漂浮瓶。



(a) YOLOv7 网络模型

(b) 改进网络模型

图 9 目标密集图片检测结果对比

Fig. 9 Comparison of detection results on an image with dense targets



(a) YOLOv7 网络模型

(b) 改进网络模型

图 10 小目标图片检测结果对比

Fig. 10 Comparison of detection results on an image with small targets

3 结论与展望

针对YOLOv7在检测河道漂浮垃圾时性能不足的情况,本研究创新性地增加了一个专门针对小目标检测的层级。这一改进显著增强了模型对小尺寸目标特征的识别和提取能力;引入Biformer注意力机制,加强特征提取能力,使模型在小目标检测上更有优越的性能;在此基础上引入MPDIoU损失函数,其具有更高的边界框回归效率和精度,同时简化计算过程,能够更好地训练模型;通过引入PELAN结构,YOLOv7模型的复杂度得到了降低,实现了检测速度和精度之间的优化平衡。实验结果显示,改进后的YOLOv7在检测性能上超越了原始模型以及其他传统的目标检测模型。

从实验结果来看,尽管改进模型的性能高于其他传统目标检测模型,但是依然存在漏检、精度较低等问题。因此,接下来的工作中,准备将数据集进行扩充,以提升对小目标的感知能力,同时考虑在此基础上对模型的轻量化进行更深入研究,以提高模型在实际应用中的检测性能。

参考文献(References):

- [1] 谷永立,宗欣欣. 基于深度学习的目标检测研究综述[J]. 现代信息科技, 2022, 6(11): 76-81.
GU Yong-li, ZONG Xin-xin. A review of object detection study based on deep learning[J]. Modern Information Technology, 2022, 6(11): 76-81.
- [2] 刘洪江,王懋,刘丽华,等. 基于深度学习的小目标检测综述[J]. 计算机工程与科学, 2021, 43(8): 1429-1442.
LIU Hong-jiang, WANG Mao, LIU Li-hua, et al. A survey of small object detection based on deep learning[J]. Computer Engineering & Science, 2021, 43(8): 1429-1442.
- [3] 潘晓英,贾凝心,穆元震,等. 小目标检测研究综述[J]. 中国图象图形学报, 2023, 28(9): 2587-2615.
PAN Xiao-ying, JIA Ning-xin, MU Yuan-zhen, et al. Survey of small object detection[J]. Journal of Image and Graphics, 2023, 28(9): 2587-2615.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6517-6525.
- [6] REDMON J, FARHADI A. Yolov3: an incremental improvement[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 89-95.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [8] 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, 2017, 39(6): 1137-1149.
- [9] ZHANG M, XU S, SONG W, et al. Lightweight underwater object detection based on YOLOv4 and multi-scale attentional feature fusion[J]. Remote Sensing, 2021, 13(22): 4706.
- [10] 戚玲珑,高建瓴. 基于改进YOLOv7的小目标检测[J]. 计算机工程, 2023, 49(1): 41-48.
QI Ling-long, GAO Jian-ling. Small object detection based on improved YOLOv7[J]. Computer Engineering, 2023, 49(1): 41-48.
- [11] DENG H, ERGU D, LIU F, et al. An embeddable algorithm for automatic garbage detection based on complex marine environment[J]. Sensors, 2021, 21(19): 6391.
- [12] ZOU P, YANG K J, LIANG C. Improving real-time detection of lightweight irregular driving behavior in YOLOv5[J]. Computer Engineering and Applications, 2023, 59(13): 186-193.
- [13] 郝琨,王阔,王贝贝. 基于改进Mobilenet-YOLOv3的轻量级水下生物检测算法[J]. 浙江大学学报(工学版), 2022, 56(8): 1622-1632.
HAO Kun, WANG Kuo, WANG Bei-bei. Lightweight underwater biological detection algorithm based on improved Mobilenet-YOLOv3[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(8): 1622-1632.
- [14] FU H, SONG G, WANG Y. Improved YOLOv4 marine target detection combined with CBAM[J]. Symmetry, 2021, 13(4): 623.
- [15] WANG C Y, BOCHKOVSKIY A, LIAO H. 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: IEEE Press, 2023: 7464-7475.
- [16] CHEN J, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 12021-12031.
- [17] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.
- [18] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1580-1589.
- [19] DO NASCIMENTO M G, PRISACARIU V, FAWCETT R. DSConv: Efficient convolution operator [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 5148-5157.

责任编辑:李翠薇