

基于YOLOv8的小目标检测方法研究

王晓雯, 常居泰, 于骥瑞

(山西大学 物理工程学院, 山西 太原 030006)

摘要: 为了提升小目标检测的准确性和效率, 本文针对现有算法在小目标识别方面的不足, 提出了一种基于YOLOv8的改进型检测方法。该方法在YOLOv8的基础上, 融入了SPD(Space-to-Depth)模块, 有效避免了传统步长卷积和池化操作所导致的信息损失; 同时, 提出了改进的FT_Conv(Fractional Fourier Transform Convolution), 提高了模型对小目标的检测精度和计算效率; 此外, 通过C2f_BiLevel Routing Attention机制, 实现了动态稀疏注意力, 优化了特征融合和物体检测性能, 进一步提升了模型对小目标的识别能力; 最后, 引入了Powerful-IoU损失函数, 以改善现有的IoU锚框面积扩大问题, 增强了对锚框的聚焦能力。实验结果表明, 相较于原始YOLOv8模型, 改进后的模型在小目标检测任务中的平均精度(AP)提升了3.29个百分点, 并且显著地降低了漏检率和误检率。这些成果证实了改进型YOLOv8模型在小目标检测领域具有明显的性能优势。

关键词: 小目标检测; YOLOv8; 分数阶傅里叶变换; 模型评估; 动态稀疏注意力; Powerful-IoU损失函数

中图分类号: TP393

文献标识码: A

doi: 10.62756/jnuc.issn.1673-3193.2025.01.0005

引用格式: 王晓雯, 常居泰, 于骥瑞. 基于YOLOv8的小目标检测方法研究[J]. 中北大学学报(自然科学版), 2025, 46(6): 712-725.

WANG Xiaowen, CHANG Jutai, YU Qirui. Research on small target detection method based on YOLOv8 [J]. Journal of North University of China(Natural Science Edition), 2025, 46(6): 712-725.

Research on Small Target Detection Method Based on YOLOv8

WANG Xiaowen, CHANG Jutai, YU Qirui

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: In order to improve the accuracy and efficiency of small target detection, this paper proposed an improved detection method based on YOLOv8 to address the shortcomings of existing algorithms in small target recognition. Based on YOLOv8, this method integrated SPD (Space-to-Depth) module, which effectively avoided the information loss caused by traditional strided convolution and pooling operation. At the same time, an improvement of Fractional Fourier Transform Convolution (FT_Conv) was proposed to improve the detection accuracy and computational efficiency of the model for small targets. In addition, the C2f_BiLevel Routing Attention mechanism was used to realize dynamic sparse attention, optimize the feature fusion and object detection performance, and further improve the recognition ability of the model for small targets. Finally, the Powerful-IoU loss function was introduced to improve the area expansion of the anchor frame of the existing IoU and enhance the focusing ability of the anchor frame. The experimental results show that compared with the original YOLOv8 model, the average accuracy

收稿日期: 2025-01-06

基金项目: 山西省基础研究计划资助项目(202403021221010)

作者简介: 王晓雯(1979-), 女, 副教授, 硕士, 主要从事网络通信方面的研究。E-mail: 381984245@qq.com.

(AP) of the improved model in small target detection tasks is increased by 3.29 percentage points, and the false detection and missed detection rates are significantly reduced. These results confirm that the improved YOLOv8 model has obvious performance advantages in the field of small target detection.

Key words: small target detection; YOLOv8; FT_Conv; model evaluation; dynamic sparse attention; Powerful-IoU loss function

0 引言

目标检测作为计算机视觉领域的核心任务之一,在自动驾驶、视频监控、医学影像分析等领域具有较高的应用价值。随着深度学习技术的快速发展,基于卷积神经网络(CNN)的目标检测算法(如 Faster R-CNN^[1]、YOLO 系列^[2])取得了显著进展,尤其是以 YOLO(You Only Look Once)系列为代表的单阶段检测模型,因其高效的检测速度成为实时目标检测的标杆。然而,尽管目标检测算法在大中型目标上表现优异,对于图像中尺寸较小的目标(通常定义为面积小于 32 像素×32 像素的目标),其检测性能仍存在显著不足。小目标由于分辨率低、特征信息匮乏,容易在检测过程中被忽视或误判^[3],这一问题在复杂场景(如遥感图像^[4]、密集人群^[5]、医学图像^[6]等)中尤为突出。因此,提升小目标检测的准确性和鲁棒性对于推动目标检测技术的实际应用具有重要意义。

为提升小目标检测性能,众多学者进行了深入的研究:Kisantal^[7]增加了基于过采样的 copy-paste(复制粘贴策略),即将小目标粘贴到图像中的任何位置,并生成新的注释,该策略通过增加小目标的数量来提高小目标检测的性能,并在实验中证明了其有效性。Akyon 等^[8]提出了一种结合切片辅助超推理(SAHI)和微调策略的创新方法,通过多尺度切片处理和针对性模型优化,显著提升了小目标检测任务中的识别精度和鲁棒性。Liu 等^[9]提出了多尺度 contourlet 知识引导学习网络,有效捕获了小尺度目标的特征。Lin 等^[10]设计的特征金字塔网络(FPN)通过跨尺度特征融合缓解尺寸差异问题。然而,传统 FPN 对小目标的浅层特征利用不足,导致细节丢失。为此,Cao 等^[11]开发了一种注意力引导的上下文特征金字塔网络(Attention-Guided Context Feature Pyramid Network),旨在解决高分辨率输入中特征图分辨率与接受域之间的矛盾。Jiang 等^[12]提出了一种名为 GiraffeDet 的新型目标检测架构,采用独特的“重颈部”设计范式,通过强化特征金字塔网络的表

征能力,显著提升了多尺度目标检测的精度和鲁棒性。此外,生成对抗网络(GAN)被用于小目标特征增强,如 Li 等^[13]提出的通过对抗学习生成高分辨率特征图。Chen 等^[14]提出了一种基于可变形 DETR 和多级特征融合的白细胞检测方法,通过改进注意力机制和特征融合策略,显著提升了血液显微图像中白细胞的检测精度,为血液疾病辅助诊断提供了有效解决方案。Xu 等^[15]提出了一种基于 Haar 小波的无参数下采样方法,用于语义分割任务,在降低计算量的同时保留了多尺度特征信息。

尽管已有研究在小目标检测领域取得了一定的进展,但随着应用场景的日益复杂,对小目标检测的准确性和鲁棒性提出了更高的要求。小目标经常在检测过程中被忽视或误判,这在一定程度上限制了目标检测技术在实际应用中的效能。因此,探索更为有效的小目标检测方法对于推动目标检测技术的发展和具有重要应用具有重要意义。本文的目的在于探索基于 YOLOv8 的小目标检测方法,旨在优化检测网络结构,增强模型对小目标特征的学习能力,从而提高检测的准确性和鲁棒性。基于此,本文提出一种改进的 YOLOv8 小目标检测算法,通过融合 SPD_Conv^[16](空间金字塔深度可分离卷积)增强多尺度特征提取,结合 Bi_Level Routing Attention^[17](双层路由注意力)实现高效区域特征筛选,并引入 FrFT(分数阶傅里叶变换)提升频域抗噪能力。相较于原始 YOLOv8 模型,改进后的模型在小目标检测任务中的平均精度(AP)有显著的提升,并且显著地降低了小目标的误检和漏检率。

1 改进的 YOLOv8 检测模型

YOLOv8 算法在目标检测中一直是研究热点,其保持了检测精度和检测速度的均衡,但在小目标检测方面,这些算法仍存在一定的局限性。小目标由于其尺寸小、特征不明显,常常在检测过程中被忽视或误判,基于此,提出了一种改进 YOLOv8 的小目标检测方法,具体工作如下:

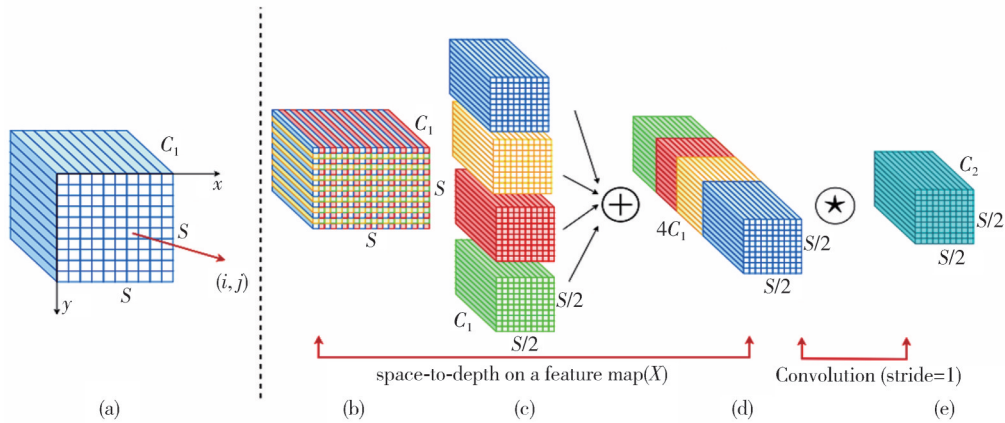


图 2 scale=2 时的 SPD_Conv 示意图

Fig. 2 Schematic diagram of SPD_Conv at scale=2

1.2 FT_Conv

分数阶傅里叶变换(FrFT)^[18]是对传统傅里叶变换的一种拓展,它突破了传统傅里叶变换只能进行整数次变换的限制,允许对信号开展非整数次的变换操作,其核心优势在于能够在时域和频域之间实现任意角度的变换。

对于给定的一个信号 $f(x)$,其 p 阶的FrFT可以通过积分形式定义为

$$F_p f(x)(u) = \int_{-\infty}^{\infty} f(x) K_p(x, u) dx, \quad (1)$$

式中: $K_p(x, u)$ 为FrFT的核函数; p 为变换的阶数。

核函数 $K_p(x, u)$ 的具体表达式为

$$K_p(x, u) = \sqrt{1 - i \cot \alpha} \cdot \exp(i\pi(x^2 \cot \alpha - 2xu \csc \alpha + u^2 \cot \alpha)), \quad (2)$$

式中: $\alpha = p \frac{\pi}{2}$ 。

FrFT具备周期性这一重要性质,具体表现为

$$F_{p+4}\{f(x)\} = F_p\{f(x)\}. \quad (3)$$

基于这一性质,在实际计算过程中,通常可以将 p 的取值范围限制在 $[0, 4)$ 这个区间内来进

行相关运算。

此外,FrFT还是一种线性变换,满足线性关系

$$F_p\{af(x) + bg(x)\} = aF_p\{f(x)\} + bF_p\{g(x)\}, \quad (4)$$

并且它是可逆的,其逆变换对应的就是 F_p 。

图 3 所示为 FT_Conv 模块的工作流程。FT_Conv 模块是一种专为小目标检测设计的卷积神经网络模块,它结合了FrFT的理论基础和传统卷积操作,通过双分支并行结构强化小目标特征提取能力:空间分支(SPU)采用多尺度卷积(3×3 和 5×5 并行)捕捉局部细节,频域分支(FPU)则利用FrFT优化高频特征表示并抑制背景干扰,从而实现对高频成分的增强。该模块通过多尺度卷积适应不同尺寸目标,并结合通道注意力和空间注意力机制动态调整特征融合的权重。同时,通过残差连接和特征拼接策略保留多尺度信息,并采用高频增强滤波器提升小目标的特征响应。在处理卷积神经网络的特征图时,FT_Conv将特征图从空间域转换到频域,通过调整变换的频率参数灵活地强调或抑制某些特征,从而提取和学习到小目标的细节和纹理特征。

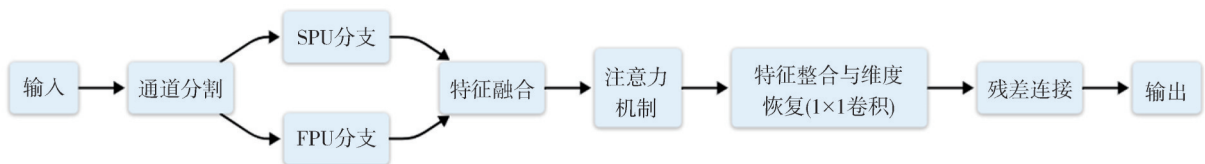


图 3 FT_Conv 模型的流程图

Fig. 3 Flowchart of the FT_Conv model

SPU 模块如图 4 所示,输入 32 通道($32 \times H \times W$)的特征图,首先通过通道分割分成两个 16 通道

数据,分别经过 3×3 卷积保留局部细节和 5×5 卷积捕获更大范围的上下文信息,再将两路结果相加,

融合多尺度信息,最后与原始输入的另外16通道数据进行拼接,恢复为32通道输出。整个模块通过

并行多尺度卷积结合残差拼接,在保持输入输出维度一致的同时,整合了多种不同尺度的特征。

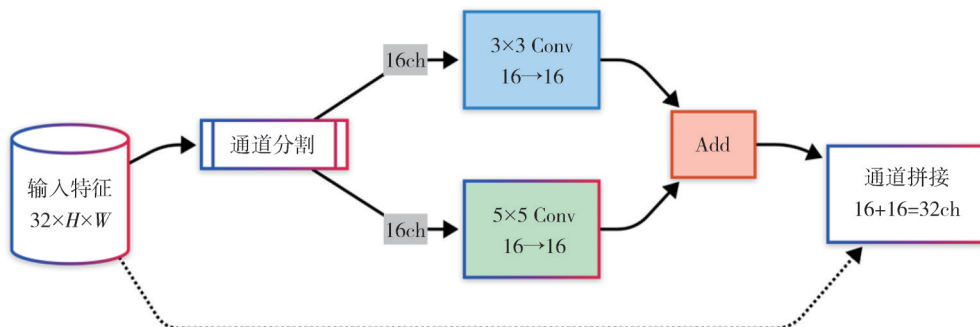


图4 SPU模块结构图

Fig. 4 Structure diagram of SPU module

FPU模块如图5所示,输入32通道($32 \times H \times W$)的特征图先被均等分割成4个8通道的子特征,分别进行不同频率参数($f=0.25/0.50/0.75/1.00$)的FrFT,通过调整变换的频率参数来

控制核函数的振荡频率,最后将处理后的4个子特征重新合并为32通道输出。该模块利用不同频率参数进行频域分解与融合,在保持特征图尺寸大小不变的前提下,提取了不同频段的信息。

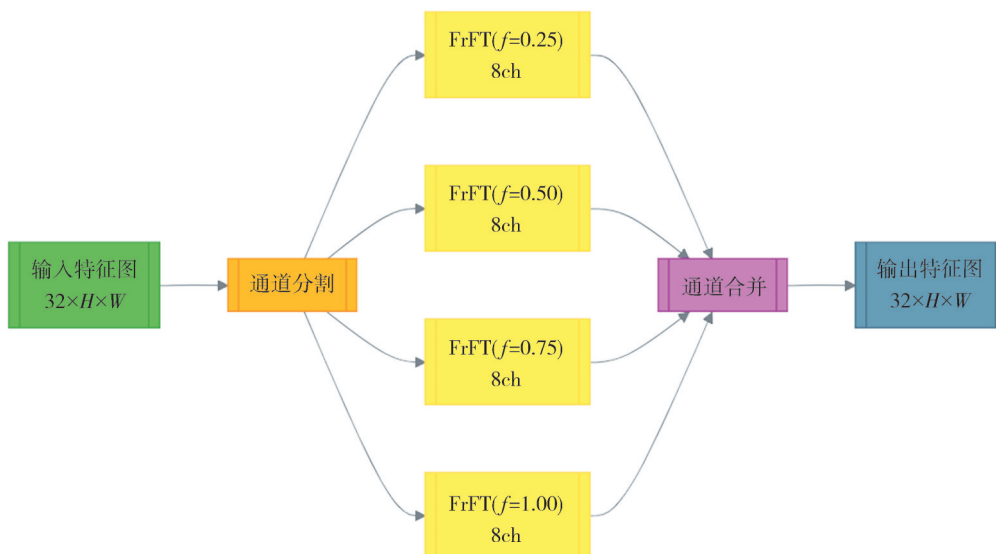


图5 FPU模块的结构图

Fig. 5 Structure diagram of FPU module

FrFT模块的卷积核由式(5)计算所得,其中, x_i 和 y_j 是网格坐标(归一化到 $[1, \text{kernel_size}]$ 区间), f 是频率参数控制核函数的频率范围,通过 $f=0.25, 0.50, 0.75, 1.00$ 生成不同频率的核,共同实现多尺度特征提取。

$$W_{i,j,f} = \cos\left(\frac{-fx_i + f^2 + x_i^2}{\sin p + 2 \tan p}\right) \cos\left(\frac{-fy_j + f^2 + y_j^2}{\sin p + 2 \tan p}\right). \quad (5)$$

1.3 C2f_BiLevel Routing Attention

小目标分辨率低、特征弱、遮挡率高,因此,

小目标检测是计算机视觉中的经典难题。C2f_BiLevel Routing Attention是BiFormer模型中的核心组件,通过动态稀疏注意力机制来解决小目标检测存在的问题,动态筛选关键区域,从而更加关注小目标所在的区域。该模型通过内容感知,只关注与查询最相关的一小部分键值对。C2f_BiLevel Routing Attention包含两个主要步骤:区域级路由(Coarse Region Level)和令牌级注意力(Token-Level Attention)。

区域级路由:首先,模型构建一个区域级关联图,通过矩阵乘法计算区域间的语义关联度。

然后,对于每个区域,模型保留与其他区域中最相关的前 k 个连接,这可通过 top- k 操作来实现。这样,每个区域只需要关注被路由到的 top- k 区域。

令牌级注意力:确定了要关注的区域后,模型在这些区域的并集中应用细粒度的令牌到令牌注意力。这一步涉及到从路由区域中收集键值对,然后进行注意力计算。

通过区域级路由和令牌级注意力的结合,C2f_BiLevel Routing Attention 保留了小目标区域,增强了局部细节(边缘/纹理),并通过双层路由来结合局部与全局信息,同时保持了对长距离

依赖的捕捉能力。

如图 6 所示,将特征图划分为 $S \times S$ 个非重叠区域,每个区域包含 $\frac{HW}{S^2}$ 个 token。对每个 token 计算 Query、Key、Value,再对每个区域的 Query 和 Key 取平均值,得到区域级表征,然后计算区域间亲和度矩阵,保留最相关的 k 个区域 Top- k ,再进行 Token 级注意力(细粒度计算),从特征图中收集 Top- k 的 Key 和 Value,对每个 Query token,仅计算已筛出的 Key/Value 的注意力,整个过程通过粗粒度路由(区域筛选)+细粒度计算(Token 级注意力)实现动态稀疏性。

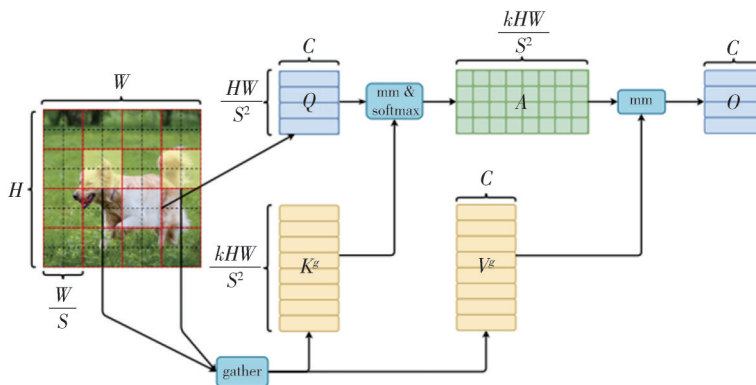


图 6 BiLevel Routing Attention 结构示意图

Fig. 6 Structure diagram of BiLevel Routing Attention

1.4 Powerful-IoU

边界框回归(Bounding Box Regression, BBR)是目标检测中的核心任务之一,而BBR的损失函数对其性能影响很大。为了解决这个问题,先对不同的IoU进行理论分析,常用的IoU包括GIoU、CIoU、EIoU以及SIoU。

图 7 所示为基于不同IoU的损失函数指导下的锚框回归过程。训练轮数(epoch)从左到右分别为20, 40和60 epochs,图中的横纵坐标分别表示归一化的坐标,用来表示target box和initial anchor box之间的相对位置关系,填充彩色的框是在回归过程中由不同损失函数引导的锚框。由图7可以看出,这些IoU引导的锚框都存在面积扩大问题,且收敛速度缓慢。分析其原因是损失函数受到不合理的惩罚因素的影响,如:使用诸如锚框和目标框的最小外部边界框的对角线长度或面积等维度信息作为损失因子的分母,虽然可以使得损失值在一个合理的范围内变化,有助于模型的收敛,也可以使损失函数对目标的尺寸变化具有一定的鲁棒性和平衡不同误差项的作用,

但是会导致锚框在回归过程中膨胀,显著减缓收敛速度。

为解决该问题,引入一种结合了目标尺寸自适应惩罚因子和基于锚框质量的梯度调节函数的Powerful-IoU (PIoU)损失函数。PIoU损耗仅使用目标框的边缘长度作为损耗因子的分母,有效地解决了面积扩大问题。PIoU损耗引导锚框沿着有效路径回归,从而可以比现有的基于IoU的损耗更快地收敛,同时引入了一个非单调注意层,将其与PIoU结合,增强了对中等质量锚框的聚焦能力。

由图7可知:以PIoU损失为导向的锚盒最快回归到接近目标盒。同时,PIoU损失引导的锚盒不存在面积扩大问题,而其他损失函数引导的锚框都存在面积扩大问题。

图8为PIoU的损失函数示意图,损失函数计算如式(6)^[19]所示。

$$L_{PIoU} = L_{IoU} + 1 - e^{-p^2}, 0 \leq L_{PIoU} \leq 2, \quad (6)$$

$$P = \frac{1}{4} \left(\frac{d\tau w_1}{\tau w_{gt}} + \frac{d\tau w_2}{\tau w_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}} \right). \quad (7)$$

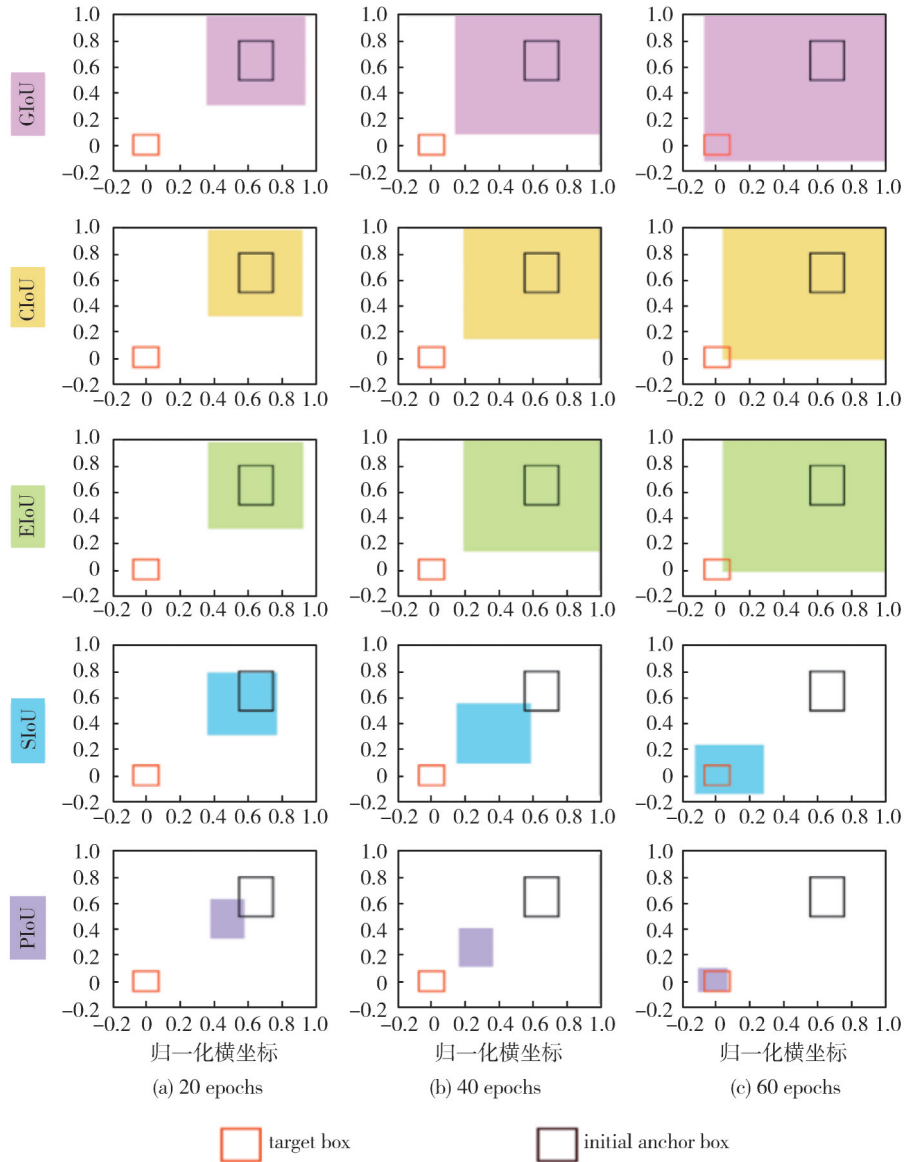


图7 不同IoU的损失函数指导下的锚框回归过程

Fig. 7 Anchor frame regression process guided by loss functions of different IoU

数据集和自制的SmallDroneDataset数据集。

2.1.1 VisDrone2019数据集

该数据集由天津大学机器学习和数据挖掘实验室 AISKYEYE 团队收集。包括288个视频片段,由261 908帧和10 209幅静态图像组成基准数据集,由多种无人机摄像头捕获,遵循了数据集的多维覆盖原则,覆盖范围广泛,包括不同位置(来自中国相隔数千公里的14个不同城市)、不同环境(城市和农村)、不同的天气(阴天、晴天、雨天等)、不同的光照条件(白天、黑夜、黄昏等)、不同物体(行人、车辆、自行车等)和不同密度(稀疏和拥挤)的场景。本数据集中一共标注了超过260万个日常目标,比如行人、汽车、自行车和三轮车。官方提供的数据中,训练集为6 471张,验证集为548张,测试集为1 610张。数

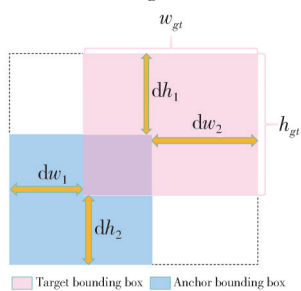


图8 PIoU损失函数示意图

Fig. 8 Schematic diagram of PIoU loss function

2 实验与分析

2.1 数据集

本文使用的数据集有两个,分别为VisDrone2019

数据集共提供了 12 个类, 分别为忽略区域、pedestrian、people、bicycle、car、van、truck、tricycle、awning-tricycle、bus、motor 和 others, 其中忽略区域和 others 是非有效目标区域, 本实验中予以忽略。

由于 VisDrone 数据集的标注文件不是 YOLO 直接可以训练的格式, 所以需进行转换。转换步骤如下: 1) 读取原始数据集中的图像文件和对应的 XML 或 JSON 标注文件; 2) 将这些标注文件中的类别信息和 bounding box 坐标转换为 YOLO 所需的文本格式, 即每个图像对应一个 .txt 文件, 文件中每行包含目标的类别索引和归一化后的中心点坐标以及宽度和高度; 3) 创建类别映射, 将类别名称映射到连续的数字 ID; 4) 根据需要调整图像尺寸以适应 YOLO 模型的输入要求; 5) 将数据集划分为训练集、验证集和测试集; 6) 生成 YOLO 配置文件, 指定数据集路径、类别数量和图像尺寸等信息, 使其能够被 YOLO 框架直接用于训练。

2.1.2 自制的 SmallDroneDataset 数据集

该数据集共包括 3 450 幅静态图像, 采用 DJI Mavic 3 Enterprise 进行拍摄, 为了确保模型在真实场景中的泛化能力, 在数据采集阶段严格遵循了多维度覆盖原则, 具体体现在: 类别多样性(包括行人、车辆、巴士以及卡车等类别), 场景复杂性(包括不同光照条件、不同天气、不同时间段以及拍摄角度等), 数据的覆盖范围广。为避免单一设备拍摄的画面对模型泛化能力的影响, 从网络搜集了其他型号无人机拍摄的画面, 共同制成数据集。然后利用 LabelImg 对感兴趣的目标框进行标注, 包括 pedestrian, people, bicycle, car, van, truck, tricycle, awning-tricycle, bus, motor。最后数据集划分为训练集 2 588 幅, 测试集 644 幅, 验证集 218 幅。



图 9 SmallDroneDataset 数据集示例图片

Fig. 9 Sample image of the SmallDroneDataset dataset

图 9 所示为 SmallDroneDataset 的部分图片, 遵循多维覆盖原则, 如左上角图片为低角度拍

摄, 右上角图片为高角度拍摄, 左下角图片存在大面积遮盖, 右下角图片为夜间拍摄, 数据覆盖维度较广。

2.2 评估指标

小目标检测算法的性能评估主要有:

1) 均值平均精度(mAP): 这是衡量小目标检测精度的核心指标, 反映了模型对图像中小目标的识别能力。mAP 值越高, 表明模型的检测精度越高, 对小目标的识别效果越好。

2) 查准率(Precision): 针对特定类别, 衡量识别出的目标中识别正确的数量占比。Precision 表示所有被检测为正例的情况下, 实际为正例的比例。

3) 召回率(Recall): 表示在所有真实正例中, 能被模型正确检测到的比例。召回率反映了模型找到所有相关案例的能力。

4) 参数量(Params): 模型所有可学习参数(如卷积核权重、全连接层参数)的总和, 用于衡量模型的复杂度, 通常以百万级或十亿级计。

2.3 实验配置

本次实验在 win11 操作系统环境下进行, 具体实验配置如下: 系统搭载了 12 个虚拟 CPU 核心, 采用 Intel Xeon Platinum 8352V 处理器, 主频为 2.10 GHz, NVIDIA RTX 3080 显卡, 20 GB 显存。实验基于深度学习框架 PyTorch 1.11.0 进行模型训练。在开始模型训练之前, 对模型配置文件进行优化设置, 包括将输入图片像素的大小统一调整为 640×640。训练过程中, 训练轮数设定为 300 轮, 批量处理大小设定为 16。选用的优化器是 SGD(随机梯度下降), 优化器动量参数设置为 0.937。初始学习率设定为 0.01, 同时, 为了正则化模型, 权重衰减系数设置为 0.000 5。在上述实验配置下完成小目标检测算法的训练。

2.4 实验结果

2.4.1 训练结果分析

本文所改进的模型在目标检测任务中表现出了卓越的性能。如图 10 所示, 模型准确地识别出了包括行人、自行车、汽车在内的多种交通参与者, 且大多数检测框与目标轮廓重叠, 证明了本文改进算法的准确性; 在目标间存在遮挡或距离较远的情况下仍然能够识别, 证明了改进后模型的鲁棒性。综上

证明本文模型在自动驾驶和交通监控等应用场景具有实际应用价值。



图10 训练结果

Fig. 10 Training results

2.4.2 热力图

图11为本文所改进的模型在小目标检测时的热力图,色彩的强度用于表示模型对行人存在的置信度。图中的高亮区域与图片中实际目标所在位置大部分重合,证明了模型在小目标检测方面的准确性,且原图片为夜间图片,光线较差,因此,该检测结果也证明了本文改进模型在复杂场景下的鲁棒性和适应性。

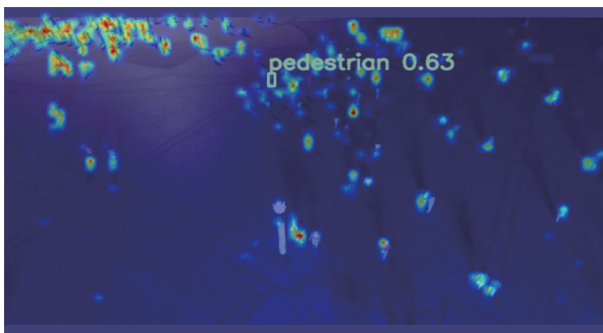


图11 热力图

Fig. 11 Heat map

2.4.3 消融实验

基于SmallDroneDataset数据集,在环境配置和训练参数一致的情况下进行消融实验,实验结果如表1所示。其中,模型A是在YOLOv8算法的基础上引入了SPD(space to depth)模块,模型B是在模型A的基础上融合了FT_Conv,模型C是

在模型B的基础上使用了C2f_BiLevel Routing Attention,模型D是在模型C的基础上将原版IoU替换为Powerful-IoU。

表1 消融实验参数对比

Tab. 1 Comparison of ablation experimental parameters

模型	$mAP_{50}/\%$	$mAP_{50-95}/\%$	$P/\%$	$R/\%$	参数量/个
YOLOv8	31.108	17.400	43.937	31.611	2.62×10^7
模型A	32.912	18.779	45.001	32.339	2.74×10^7
模型B	32.656	18.874	41.134	34.226	3.02×10^7
模型C	33.644	19.33	44.176	33.485	2.56×10^7
模型D	34.395	19.695	48.773	32.638	2.56×10^7

由表1可以看出,相比YOLOv8,模型A的 mAP_{50} 值上升了1.804个百分点, mAP_{50-95} 值上升了1.379个百分点,准确率和召回率分别提升了1.064个百分点和0.728个百分点。以上结果表明,在YOLOv8算法的主干部分添加SPD_Conv有效地避免了传统的步长卷积和池化过程中的信息丢失,可以提高算法对小物体和低分辨率图像的处理能力;相比模型A,模型B的 mAP_{50-95} 值提升0.095个百分点,而参数量只增加了 2.8×10^6 个,表明通过优化的频域处理,可以在不显著增加计算负担的情况下提升性能;模型C中用C2f_BiLevel Routing Attention代替了原版的C2f,相比模型B,其 mAP_{50} 和 mAP_{50-95} 指标分别提升了0.988个百分点和0.456个百分点,准确率提升了3.042个百分点,且参数量减少了 4.6×10^6 个,在减轻计算负担和内

存压力的同时实现了更灵活的计算分配和内容感知;模型D用Powerful-IoU替换了原版的IoU,相较于模型C,其 mAP_{50} 和 mAP_{50-95} 指标分别提升了0.751个百分点和0.365个百分点,准确率提高了4.597个百分点。最终,改进版YOLOv8比原版YOLOv8的 mAP_{50} 和 mAP_{50-95} 指标分别提升了3.287个百分点和2.295个百分点,准确率和召回率分别提升了4.836个百分点和1.027个百分点,证明改进YOLOv8算法的小目标检测性能有一定的提升。

由图12可以看出,所有模型的精确率随着训练轮数的增加而提高,最终趋于稳定。模型C的

精确率在大部分训练过程中高于其他模型;各模型的召回率同样随着训练轮数的增加而提高。模型C在召回率上也同样表现优异,显示出了较好的性能;模型C在 $mAP@IoU=0.5$ 指标上的表现优于其他模型,其 mAP_{50} 值明显高于模型A、模型B和YOLOv8;模型C在 mAP_{50-95} 这个更严格的指标上也显示出了较好的性能,其曲线在训练过程中始终位于其他模型之上。总体来说,模型C在所有性能指标上都显示出了较好的性能,这表明模型C具有较好的学习能力和收敛性,在这次实验中的整体表现更优,证明了改进的有效性。

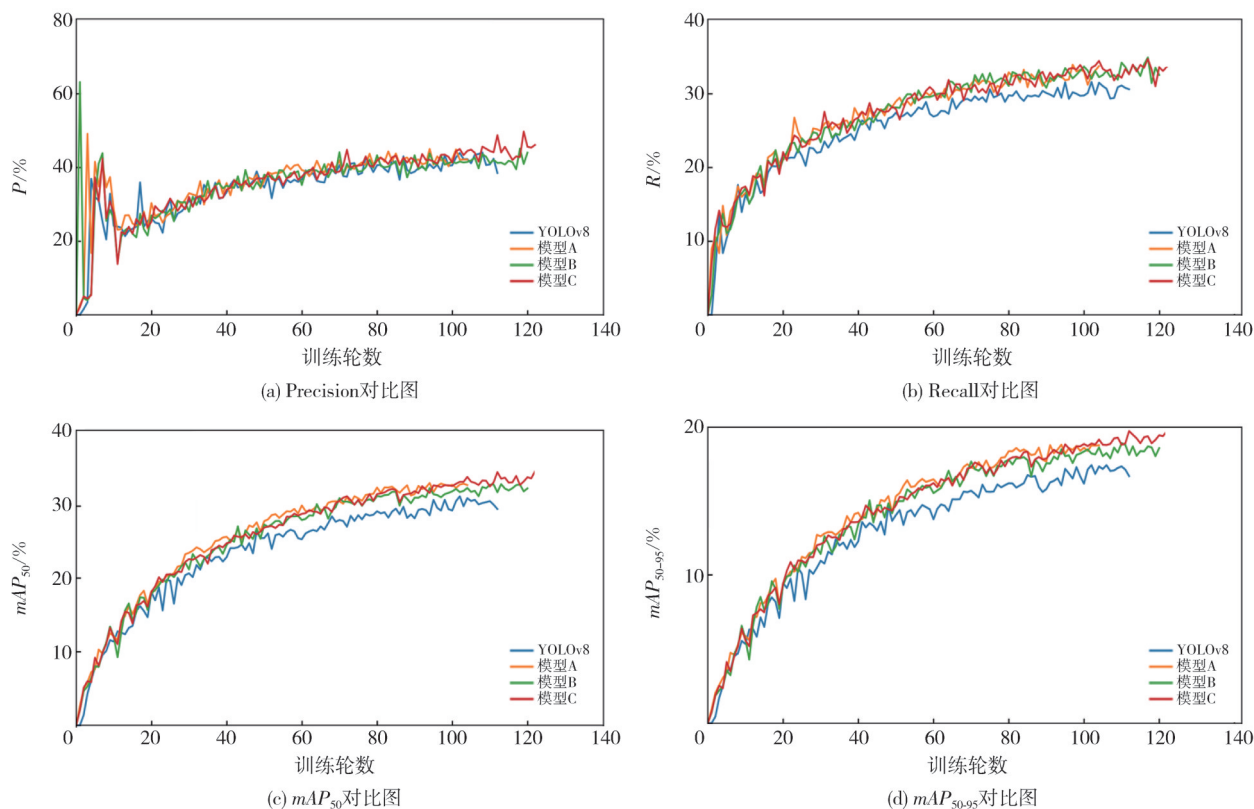


图12 各模型的性能曲线对比图

Fig. 12 Comparison of performance curves of each model

表2为不同模型在SmallDroneDataset数据集上的实验结果。模型A是在YOLOv8中融合了SPD(space-to-depth)模块;模型E是在YOLOv8的基础上结合了FT_Conv;模型F是在YOLOv8的基础上引入了C2f_BiLevel Routing Attention来代替原版的C2f。

由表2可以看出,相比YOLOv8,模型E的 mAP_{50} 提升了0.71个百分点, mAP_{50-95} 提升了0.829个百分点,参数量增加了 2.4×10^6 个,以上结果证明,FT_Conv通过结合分数阶傅里叶变换和通道注意力机制,增强了模型对多尺度和多方向

特征的捕捉能力,从而提高了目标检测的准确性。这种设计有助于模型在保持计算效率的同时提升对小目标和复杂背景中目标的检测性能。相比YOLOv8,模型F的 mAP_{50} 提升了0.794个百分点, mAP_{50-95} 提升了0.892个百分点,参数量减少了 5×10^6 个。以上结果说明,YOLOv8结合C2f_BiLevel Routing Attention,通过动态稀疏注意力和双层路由机制优化了计算分配和内容感知能力。

由图13可以看出,相比YOLOv8,模型A、模型E和模型F的各项性能指标均有显著提升。

在准确率方面,模型A、模型E和模型F的曲线均略高于YOLOv8,证明它们在识别正类样本方面具有微弱优势。模型A在召回率上的表现较为突出,其曲线显著高于YOLOv8,这表明模型A捕捉正类样本的性能有显著提升,而模型E和模型F则在召回率上略有提升。在 $mAP@IoU=0.5$ 指标上,模型E和模型F的曲线同样略高于YOLOv8,这在一定程度上反映了它们的平均精度有轻微提升;模型A的曲线则显著高于YOLOv8,这不仅证明了其在平均精度上的显著优势,也突显了其在模型改进上的有效性。在更为严格的 mAP_{50-95}

指标下,模型E和模型F的曲线略高于YOLOv8,显示了它们更广泛IoU阈值范围内的性能提升,而模型A的曲线则再次显著超越了YOLOv8。

表2 改进模型与YOLOv8的参数对比

Tab. 2 Comparison of parameters between improvement model and YOLOv8

模型	$mAP_{50}/\%$	$mAP_{50-95}/\%$	$P/\%$	$R/\%$	参数量/个
YOLOv8	31.108	17.400	43.937	31.611	2.62×10^7
模型A	32.912	18.779	45.001	32.339	2.74×10^7
模型E	31.818	18.229	42.995	30.458	2.86×10^7
模型F	31.902	18.292	43.850	31.134	2.12×10^7

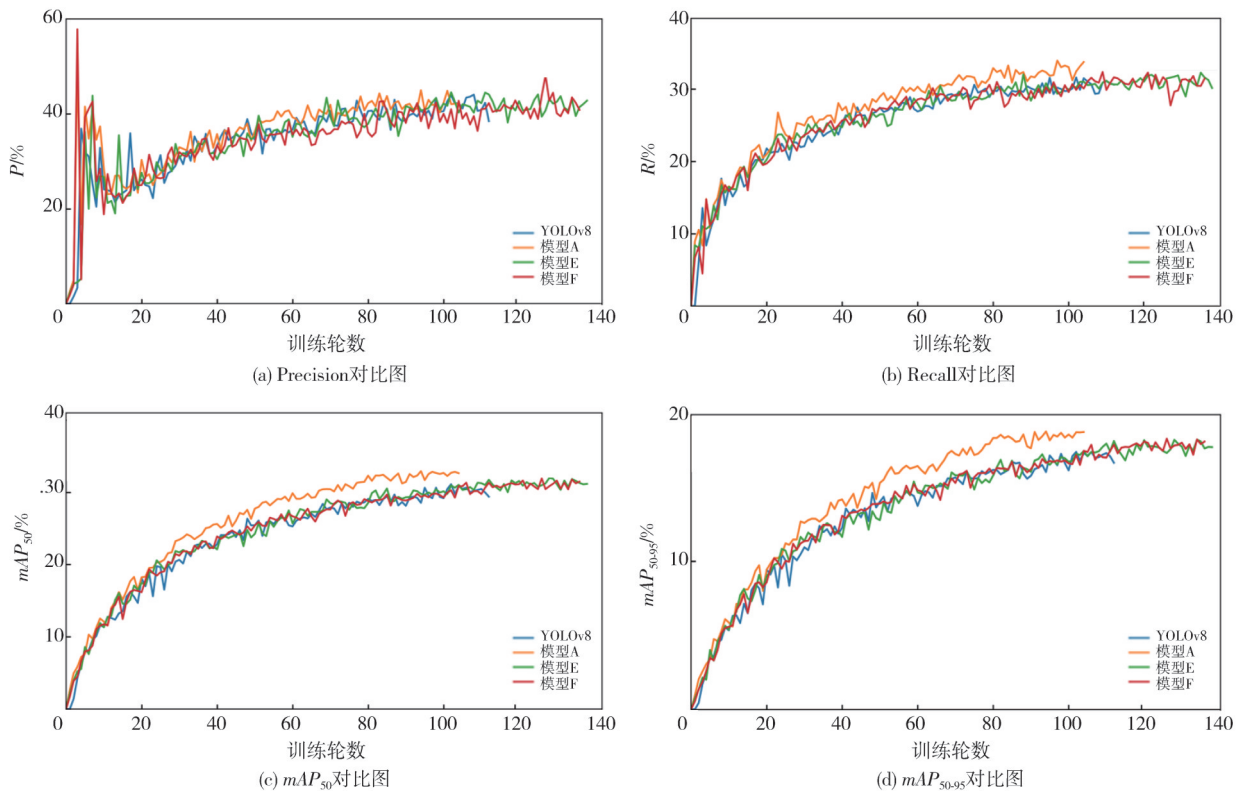


图13 模块性能对比图

Fig. 13 Performance comparison of individual modules

综合上述指标可以得出结论,模型A在所有测试的性能指标上均展现出了显著的改进效果,而模型E和模型F虽然提升幅度较小,但也在大多数指标上显示出了较好的改进效果,证明了模型改进的有效性。

2.4.4 Powerful-IoU模块性能的验证实验

为了验证不同IoU对模块性能的影响,在SmallDroneDataset数据集上开展了实验,结果如图14所示。模型D是在模型C中引入Powerful-IoU,模型G是在模型C中引入Inner-IoU^[20],模型H是在模型C中引入Shape-IoU^[21],模型I是

在模型C引入Shape_Inner-IoU。

由图14可以看出,虽然模型D持续训练至120轮左右而其他模型(G/H/I)在90轮左右时就停止,但这是正常现象。主要原因是不同模型的收敛速度存在自然差异:模型D的各项指标在90轮后仍保持稳定提升趋势,说明仍在有效学习阶段,此时其他模型的评估指标已进入平台期,触发了早停机制自动终止训练。这种差异完全符合深度学习训练规律,所有模型都达到了各自的性能最优状态,实验数据真实可靠,不影响结果的有效性和可比性。

模型D在多个关键性能指标上展现出了卓越的性能。在准确率(Precision)方面,模型D的曲线显著高于其他模型,这表明其在预测正类样本时的准确性有显著提升。在召回率(Recall)指标上,模型D同样表现突出,其曲线的优越性证明

了其在识别所有正类样本方面的高效率。此外,在 mAP_{50} (平均精度均值在IoU阈值为0.5时)和 mAP_{50-95} (平均精度均值在IoU阈值为0.5~0.95时)这两个衡量模型整体性能的严格指标上,模型D的曲线均显示出显著的领先优势。

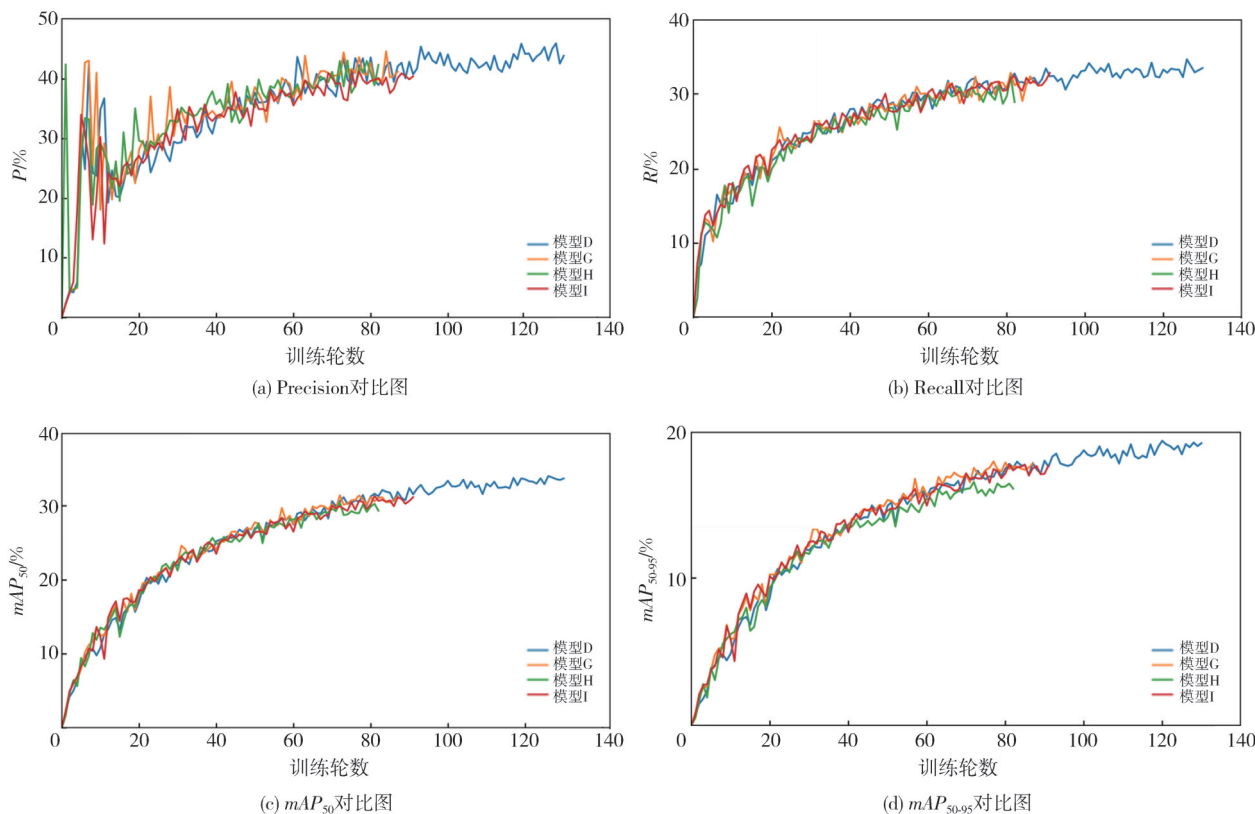


图 14 不同IoU算法的性能对比

Fig. 14 Performance comparison of different IoU algorithms

这些性能指标的显著提升证明了 Powerful-IoU 在小目标检测中的有效性。Powerful-IoU 作为一种改进的损失函数,通过更加精确地衡量预测框与真实框之间的重叠程度,可以使模型更好地学习和优化,从而能在小目标检测任务中实现更高的准确率和召回率。这一点在模型D的性能表现中得到了明确的体现,验证了 Powerful-IoU 的有效性。

2.4.5 不同算法的对比实验

为了进一步验证本文所提出的模型在小目标检测中的有效性,将改进后的YOLOv8与目前主流的目标检测算法在VisDrone2019数据集上进行了对比实验,其中包括一些无锚框的算法,默认仅计算 mAP 指标,为了增强比较的全面性和普适性,仍然将相关数据列入表中,空白数据用“—”表示,实验结果见表3。

由表3可以看出,改进后的YOLOv8模型在

目标检测任务中表现出色,其精确率(P)达到了52.627%,召回率(R)达到了41.915%。IoU为0.5时的平均精度(mAP_{50})为42.397%,而IoU为0.5~0.95的平均精度(mAP_{50-95})为25.707%。与YOLOv5系列模型相比,改进YOLOv8的性能有明显提升。YOLOv5系列的模型,n、s、m、L、x模型性能逐步增强,但改进YOLOv8性能相较于YOLOv5x,在 mAP_{50} 和 mAP_{50-95} 两项指标中,分别有0.012和0.01的提升;与YOLOv4相比,改进YOLOv8的 mAP_{50} 和 mAP_{50-95} 也显示出更好的性能。与其他一些模型如Faster R-CNN、Retina Net等相比,改进YOLOv8的性能仍具有优势,这进一步证明了改进YOLOv8在目标检测任务中的优越性能。综上所述,改进YOLOv8模型在目标检测的关键性能指标上超越了许多现有的流行模型,证明了其在小目标检测领域的实用性和有效性。

表3 不同算法的参数对比

Tab. 3 Comparison of parameters of different algorithms

模型	P/%	R/%	$mAP_{50}/\%$	$mAP_{50-95}/\%$
YOLOv5n	39.4	29.3	27.8	14.1
YOLOv5s	48.8	35.3	35.8	19.8
YOLOv5m	50.6	37.8	38.4	22.2
YOLOv5L	53.0	39.7	40.3	23.2
YOLOv5x	55.2	40.5	41.2	24.7
YOLOv4	43.4	35.1	36.3	21.4
YOLOX_L	—	—	38.8	22.2
YOLOX_M	—	—	36.6	20.7
YOLOX_S	—	—	32.6	17.9
YOLOX_Tiny	—	—	19.9	10.0
YOLOX_Nano	—	—	13.6	6.7
Faster_RCNN	—	—	37.1	21.9
Retina Net	—	—	19.1	10.6
cascade	—	—	39.0	24.5
VF Net	—	—	37.3	23.1
TOOD	—	—	39.8	24.4
DDOD	—	—	38.2	23.3
FCOS	—	—	31.9	19.0
Center Net	—	—	33.6	18.7
ATSS	—	—	36.3	22.1
FSAF	—	—	36.4	20.8
YOLOv11s	52.8	39.1	40.8	24.5
YOLOv10s	51.9	38.9	40.7	24.5
YOLOv9s	52.0	39.5	41.0	24.7
YOLOv3 Tiny	37.9	24.2	23.3	12.8
YOLOv6s	—	—	34.3	20.3
YOLOv6s mbla	—	—	36.1	21.1
改进 YOLOv8 (本文)	52.627	41.915	42.397	25.707

3 结论

本文提出了一种改进的YOLOv8的小目标检测方法,并在多个数据集上进行了实验。在Vis-Drone2019数据集上的实验结果表明,本文改进的YOLOv8算法将准确率、召回率、 mAP_{50} 和 mAP_{50-95} 分别提升至52.627%、41.915%、42.397%和25.707%,性能相较于其他流行模型仍具有性能优势,这证明了改进模型的有效性。在SmallDroneDataset数据集上的试验结果表明,相较于YOLOv8,改进YOLOv8算法的准确率、召回率、 mAP_{50} 和 mAP_{50-95} 等指标分别提升了4.836百分点、1.027百分点、3.287百分点和2.295百分点,并能够有效解决小目标检测过程中小目标的漏检问题和特征近似的目标的误检问题,模型优势明显,证明了改进的YOLOv8算法的泛化能力和有效性。

参考文献:

[1] REN S, HE K, GIRSHICK R, et al. Faster RCNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern

Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[DB/OL]. (2015-05-09)[2025-01-06]. <https://arxiv.org/abs/1506.02640>.
- [3] YANG J, WANG T. Small object detection model for remote sensing images combining super-resolution assisted reasoning and dynamic feature fusion[J]. Journal of Applied Remote Sensing, 2024, 18(2): 028503.
- [4] YIN Q, HU Q, LIU H, et al. Detecting and tracking small and dense moving objects in satellite videos: A benchmark[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5612518.
- [5] ZHANG Y, ZHAO H, DUAN Z, et al. Congested crowd counting via adaptive multi-scale context learning[J]. Sensors, 2021, 21(11): 3777.
- [6] JIANG H, DIAO Z, SHI T, et al. A review of deep learning-based multiple-lesion recognition from medical images: Classification, detection and segmentation [J]. Computers in Biology and Medicine, 2023, 157: 106726.
- [7] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection [DB/OL]. (2019-02-19)[2025-01-06]. <https://arxiv.org/abs/1902.07296>.
- [8] AKYON F C, ONUR ALTINUC S, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection[C]//2022 IEEE International Conference on Image Processing (ICIP), 2022: 966-970.
- [9] LIU M, JIAO L, LIU X, et al. Multi-scale contourlet knowledge guide learning segmentation[J]. IEEE Transactions on Multimedia, 2024, 26: 4831-4845.
- [10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936-944.
- [11] CAO J, CHEN Q, GUO J, et al. Attention-guided context feature pyramid network for object detection [DB/OL]. (2020-05-23)[2025-01-06]. <https://arxiv.org/abs/2005.11475>.
- [12] JIANG Y, TAN Z, WANG J, et al. GiraffeDet: A heavy-neck paradigm for object detection [DB/OL]. (2022-02-09)[2025-01-06]. <https://arxiv.org/abs/2202.04256>.
- [13] LI J, LIANG X, WEI Y, et al. Perceptual generative adversarial networks for small object detection [C]//

- 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1951-1959.
- [14] CHEN Y, ZHANG C, CHEN B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases [J]. *Computers in Biology and Medicine*, 2024, 170: 107917.
- [15] XU G, LIAO W, ZHANG X, et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation [J]. *Pattern Recognition*, 2023, 143: 109819.
- [16] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [DB/OL]. (2022-08-07) [2025-01-06]. <https://arxiv.org/abs/2208.03641>.
- [17] ZHU L, WANG X, KE Z, et al. BiFormer: Vision transformer with bi-level routing attention [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 10323-10333.
- [18] 陶然, 齐林, 王越. 分数阶 Fourier 变换的原理与应用 [M]. 北京: 清华大学出版社, 2004.
- [19] LIU C, WANG K, LI Q, et al. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism [J]. *Neural Networks*, 2024, 170: 276-284.
- [20] ZHANG H, XU C, ZHANG S J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box [DB/OL]. (2023-11-14) [2025-01-06]. <https://arxiv.org/abs/2311.02877>.
- [21] ZHANG H, ZHANG S J. Shape-IoU: More accurate metric considering bounding box shape and scale [DB/OL]. (2023-12-29) [2025-01-06]. <https://arxiv.org/abs/2312.17663>.