

面向无人机视角的小目标检测模型

林昊, 徐建宇, 杨欢

(华南师范大学软件学院, 广东佛山 528225)

摘要: 为解决无人机航拍图像目标检测中小目标检测精度低下的问题, 提出一种基于改进YOLOv5s的无人机航拍图像小目标检测算法。首先, 为提高主干网络的特征提取能力, 以提取更多小目标的语义信息, 将主干网络中的C3模块引入密集连接思想, 改进为DC3模块; 其次, 将CIOU损失函数替换为EIOU损失函数, 进一步加快模型收敛速度, 提升模型检测能力及定位精度; 最后, 增加一个额外的小目标检测头, 显著提升对小目标的检测精度。在无人机航拍图像VisDrone2019数据集上的实验表明, 改进后的模型与原始YOLOv5s相比, mAP@0.5提高了6.6个百分点, mAP@0.5:0.95提高了4个百分点, 更适用于无人机航拍图像的目标检测。

关键词: 无人机; 小目标检测; YOLOv5s; 密集连接; EIOU

DOI: 10.11907/rjdk.241489

中图分类号: TP183; TP391.41

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)008-0182-06



A Small Object Detection Model for Drone Perspective

LIN Hao, XU Jianyu, YANG Huan

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: To address the issue of low detection accuracy for small targets in UAV aerial imagery, an improved YOLOv5s algorithm for small target detection in UAV aerial imagery is proposed. Firstly, to enhance the feature extraction capability of the backbone network and thus capture more semantic information of small targets, the idea of dense connections is introduced into the C3 module of the backbone network, which is improved into the DC3 module. Secondly, the CIOU loss function is replaced with the EIOU loss function to further accelerate the convergence speed of the model and enhance its detection capabilities and localization accuracy. Finally, an additional small target detection head is added, which greatly improves the detection accuracy for small targets. Experiments on the VisDrone2019 dataset of UAV aerial imagery demonstrate that the improved model, compared to the original YOLOv5s, has increased the mAP@0.5 by 6.6 percentage points and mAP@0.5:0.95 by 4 percentage points, making it more suitable for target detection in UAV aerial imagery.

Key Words: unmanned aerial vehicle; small object detection; YOLOv5s; dense connection; EIOU

0 引言

在微电子软硬件的持续演进、导航与通信技术的不断革新, 以及材料和能源技术的显著突破的背景下, 无人机平台已在全球范围内成为遥感及目标检测领域备受瞩目的研究对象, 并呈现出迅猛的发展态势。与传统的信息获取平台相比, 无人机弥补了因气候、空间等限制造成的信息缺失, 且因其操作灵活、成本低廉、数据获取高效等特

点, 在土地覆盖测绘^[1]、智慧农业^[2]、智慧城市^[3]等多个领域得到广泛应用。无人机提供了一个全新的视角, 使人们能够获取到以往难以获取的地面信息。然而, 如何从大量的航拍图像中快速、准确地提取有用信息, 成为一个亟待解决的问题。无人机航拍图像的目标检测技术正是为解决这一问题而提出。目标检测作为计算机视觉领域的一个重要研究课题, 在实际应用中发挥了极大的作用。在无人机航拍图像的应用背景下, 目标检测技术不仅能够提高数据处理的效率, 还能为决策提供科学依据。

收稿日期: 2024-05-28

扫描二维码阅读全文:



基金项目: 广东省基础与应用基础研究基金项目(2020A1515110783)

作者简介: 林昊(1999-), 男, 华南师范大学软件学院硕士研究生, 研究方向为目标检测; 徐建宇(2001-), 男, 华南师范大学软件学院硕士研究生, 研究方向为目标检测; 杨欢(1986-), 男, 博士, 华南师范大学软件学院副研究员、硕士生导师, 研究方向为信息物理系统。本文通讯作者: 杨欢。

目标检测算法主要分为两阶段模型和单阶段模型两大类。其中,两阶段目标检测算法首先生成一系列候选区域,然后对这些区域进行分类和边界框回归。这类算法通常在准确性上表现更好,但速度相对较慢,代表算法包括R-CNN^[4]及其后续改进的Faster R-CNN^[5]等模型。而单阶段目标检测算法的主要特点是直接在输入图像上预测目标的类别和位置,无需先生成候选区域。这类算法通常更快、更简洁,适合实时应用场景,代表算法有YOLO系列^[6]、SSD^[7]、RetinaNet^[8]等。

尽管目标检测技术已取得显著进展,但从无人机应用的角度来看,仍面临诸多挑战与困难。在执行航空摄影任务时,无人机可能会面临多种复杂的环境挑战,如多变的天气条件、光照变化以及障碍物干扰等,这些因素均可能对图像质量和目标检测的准确性造成影响。例如,雨雪天气可能会导致图像模糊,而建筑物和树木的遮挡可能会使目标部分不可见。此外,航空图像中的目标物体尺度变化较大,从小型物体(如个人物品)到大型目标(如建筑物)均可能出现,这要求对象检测模型能够处理不同尺度的目标,且在目标尺寸变化时也保持高检测精度。同时,无人机通常携带有限的计算资源,这需要物体检测模型足够高效,以实现实时或接近实时的处理。然而,高精度的物体检测模型往往需要大量的计算资源,这对无人机平台来说是一个挑战。

在无人机航拍图像的目标检测以及小目标检测领域,国内外的学者们展开了一系列的研究。无人机的飞行高度导致目标在图像中呈现尺度不一、尺寸较小以及排列密集等特点,这些因素限制了可获取的有效特征信息数量。因此,许多研究工作聚焦于小目标检测问题,提出了多种改进方法,例如RRNet^[9]、HRDNet^[10]、Cascade网络^[11]等。Chen等^[12]提出了一种ResNeXtd组合结构,以增强对小尺寸物体的感知。Zhu等^[13]提出了TPH-YOLOv5,通过引入Transformer Prediction Heads (TPH)和CBAM注意力来增强模型对不同尺度目标的检测能力,特别是在高密度场景中的性能。Sunkara和Luo提出了SPD-Conv^[14],用以替代步幅卷积和池化层,缓解了细粒度信息的丢失问题,在小目标和低分辨率图像上显著提高了检测和分类的准确性。苏佳等^[15]通过在YOLOv5中引入自适应特征提取模块(AFE)、三重特征融合机制(TFF)和输出重构模块(ORS),有效地提升了小目标的检测精度和速度。李国强等^[16]通过将SSD的特征提取网络替换为Resnet-50,并采用特征融合技术构建特征金字塔,还设计了一种联合注意力机制的多尺度卷积结构,增强了对不同尺寸物体的检测能力。吴明杰等^[17]在YOLOv5中引入双层路由注意力和结合注意力机制的动态目标检测头,并采用Focal-EIoU损失函数解决回归预测中的不准确性问题。

针对当前方法存在的不足,本文对YOLOv5s进行了改进,主要贡献如下:

(1)通过将密集连接的思想引入C3模块,提出一个全新的DC3模块。这项改进缓解了消失梯度问题,加强了特征传播,并显著提升了主干网络的特征提取能力,从而进一步提高检测精度。

(2)EIoU损失函数对目标框的定位精度更加敏感,可以促使模型更准确地预测目标框的位置和大小,提高目标检测任务的定位精度。本文将CIoU损失函数替换为效果更好的EIoU,使模型收敛速度更快,进一步提升精度。

(3)针对数据集特点,优化检测头结构,添加小目标检测头,使模型可以更好地处理不同尺寸的目标,增强了模型对各种目标尺寸的泛化能力,大幅提升模型的检测精度。

1 YOLOv5s算法原理

YOLOv5(You Only Look Once version 5)是一种单阶段目标检测算法,其核心优势在于采用了端到端的检测框架。该算法将目标检测任务视为回归问题,通过单一的神经网络模型同时实现目标的定位和分类。这样的设计使得YOLOv5在处理图像时能够一次性给出所有目标的预测,极大地提高了检测的速度和效率,适用于实时目标检测场景,同时也相对符合无人机航拍场景的目标检测算法之一。此外,YOLOv5有n、s、m、l、x五种尺寸规格,且其网络结构相同。但是对应不同的网络深度与宽度,较小的尺寸可以减少模型的存储需求和计算开销,深层网络则可提取更丰富特征。综合考虑检测的速度和精度,本文决定选择YOLOv5s作为基准模型进行改进。

在YOLOv5的设计中,数据增强技术得到了创新的应用。通过Mosaic数据增强,算法在训练过程中将4张不同的图像拼接成一张,从而在单次前向传播中模拟更多样的场景,有效提升了模型对各种情况的适应能力和鲁棒性^[18]。此外,自适应锚框计算和自适应图片缩放技术的引入,使YOLOv5能够更好地适应不同尺寸和比例的目标,进一步提升检测的准确性。

YOLOv5的结构可清晰划分为3个核心组件:骨干网络(Backbone)专注于特征提取,颈部网络(Neck)实现特征融合,而头部网络(Head)则承担检测任务。三者紧密协作,共同构成了YOLOv5的完整架构。

骨干网络部分主要由Conv、C3等一系列模块组成。其中,C3模块用于从先前的输入中提取特征,其执行次数与网络性能直接相关。这些结构通过高效的特征提取和跨阶段的特征融合,提高了模型对目标特征的捕捉能力。在颈部网络部分,是FPN+PAN结构的结合,使得模型在处理不同尺度的特征时更为灵活,增强了对小目标和复杂背景的检测能力。在检测头部分,3个检测头被分别用来检测大、中、小3个不同类型的目标,这样的设计使得模型对于不同大小的目标都有一定的应对能力。

在实际应用中,YOLOv5展现出了强大的性能。无论是在实时视频监控、自动驾驶车辆的感知系统,还是在工业自动化与智能交通管理等领域,YOLOv5都能提供高效且可靠的目标检测服务。

2 本文算法

本文选用 YOLOv5s 作为基线模型进行改进。YOLOv5s 模型具有参数量少、检测速度快特点,并且在工业领域应用广泛,因此适合用于本文的算法研究。首先,提出一个改进后的 C3 模块,命名为 DC3 模块,并将主干网络中的 C3 模块替换成 DC3 模块;其次,使用 EIOU 损失函数替换原本的 CIOU 损失函数;最后,在检测头部分加入小目标检测头。改进后的模型网络结构如图 1 所示。

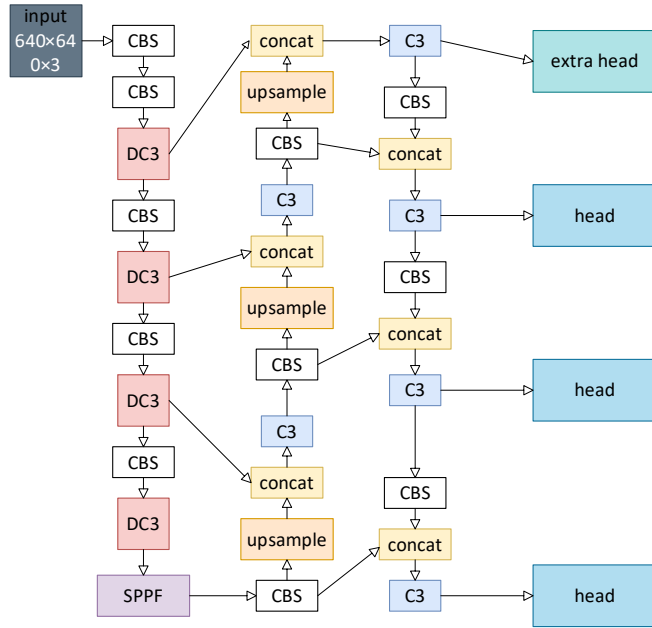


Fig. 1 Improved YOLOv5s model structure

图1 改进的 YOLOv5s 模型结构

2.1 改进的 C3 模块

在传统的卷积网络中,第 l 层的输出是下一层 $(l+1)$ 的输入,可以表示为 $\chi_l = H_l(\chi_{l-1})$ 。其中, χ_l 是第 l 层的输出, H_l 是第 l 层的转换函数, χ_{l-1} 是第 $l-1$ 层的输出。在密集卷积网络中,可以表示为 $\chi_l = H_l([\chi_0, \chi_1, \dots, \chi_{l-1}])$ 。其中 $[\chi_0, \chi_1, \dots, \chi_{l-1}]$ 是一个拼接操作,表示将前面所有层的特征图拼接在一起作为第 l 层的输入。

密集卷积网络(DenseNet),通过前馈方式将每一层连接到所以其他层。传统的卷积网络有 L 层,每层和后续层之间有一个连接,而密集卷积网络有 $L(L+1)/2$ 个直接连接。对于每个层,所有先前层的特征图被用作输入,同时其自身的特征图被用作所有后续层的输入^[19]。

受到密集卷积网络的启发,本文将密集连接的思想引入 C3 模块,将改进后的新模块命名为 DC3 模块。如图 2 所示,为充分发挥密集连接的优势,本文在 C3 模块中额外增

加了 2 个 ConvBnSilu 层。对于这 5 个 ConvBnSilu 层,本文将每个层所有先前层的特征图用作输入,同时将其自身的特征图被用作所有后续层的输入。这里的密集连接采用合并操作,即在 channel 维度上进行拼接。通过引入密集连接,特征的重用率得到提升,使得主干网络在不增加太多参数量的情况下加深其深度。这个改进缓解了消失梯度问题,增强了特征传播和主干网络的特征提取能力,从而在一定程度上提高了检测精度。

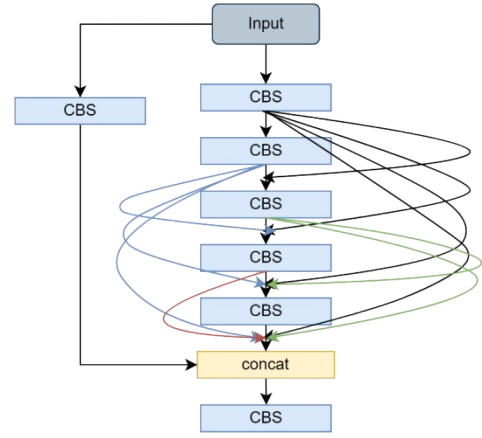


Fig. 2 DC3 module structure

图2 DC3 模块结构

2.2 损失函数的改进

在深度学习模型的训练过程中,损失函数起着至关重要的作用,其主要功能是指导模型的训练和评估,通过调整参数以更好地拟合训练数据并提高性能。YOLOv5 模型的损失函数由三部分组成:边界框回归损失、置信度损失和分类损失。其中,边界框回归损失函数采用了 CIOU。具体公式如下:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (1)$$

$$\alpha = \frac{v}{(1 - IOU) + v} \quad (2)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \quad (3)$$

$$IOU = \frac{A \cap B}{A \cup B} \quad (4)$$

式中: α 是平衡比例的参数; v 是衡量长宽比一致性的参数; b 代表预测框的中心点; b^{gt} 代表真实框的中心点; ρ 代表两个中心点之间的欧氏距离; c 代表真实框和预测框的最小外接框的对角线距离; w 和 h 分别代表预测框的宽度和高度; w^{gt} 和 h^{gt} 分别代表真实框的宽度和高度。

虽然 CIOU 损失函数考虑了边界框回归的重叠面积、中心点距离和纵横比,但是在其公式中所体现的仅仅是纵横比之间的差异,而非宽度、高度分别与其置信度之间的真实差异。具体而言,在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, 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。然而,当预测框与真实框的长宽比一致但长宽值不相等时,CIOU损失函数会失效,导致不利于损失函数的充分表达,从而限制模型的收敛速度和准确性。针对这个问题,本文将CIOU替换为EIOU,EIOU损失函数包括三个部分:重叠损失、中心距离损失和宽高损失^[20]。前两部分使预测框和真实框的宽度和高度之差最小,第三部分的宽高损失使目标框与锚框的宽度和高度值最小,从而加速收敛速度。EIOU的公式如下:

$$L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w)}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (5)$$

其中, C_w 和 C_h 分别代表真实框及预测框的最小外接框的宽和高。

2.3 检测头的改进

YOLOv5的原始模型架构包含3个不同尺度的检测头。输入图像经过主干网络的不同倍数下采样后,通过颈部网络特征融合,生成80×80、40×40和20×20三种不同分辨率的特征图。这些特征图分别被分配到3个检测头,用于执行小目标、中目标和大目标的检测任务。这样的结构使得YOLOv5能够应对常规图像中不同大小目标的检测。

然而,无人机视角下的广阔视野在提供更为丰富可视化信息的同时,也意味着更为复杂的场景和更加多样的目标。其成像视角不同于常规图像,尤其是在较高飞行高度下,图像中的小目标尺寸相较于常规图像更小,且目标分布更为密集。因此,YOLOv5原来的3个检测头的结构在检测无人机航拍图像时效果难以令人满意,对微小目标容易出现漏检的情况。为解决这一问题,本文在YOLOv5的检测头结构中引入了一个额外的微小目标检测头,用于检测分辨率更高(160×160)的特征图,使其能够保留更多小目标的语义信息。此外,本文增加的检测头采用更小的锚框,以适配小目标的尺寸范围。通过这一改进,可以显著提升小目标的检测精度。

3 实验分析

3.1 数据集

本文采用VisDrone2019数据集^[21],该数据集是由天津大学机器学习和数据挖掘实验室AISKYEYE团队构建的大型基准数据集,旨在促进计算机视觉领域与无人机技术相结合的研究发展。其中包含288个视频片段和10 209张图像,这些数据由部署在中国14个不同城市的无人机摄像头采集,涵盖了多样化的环境和场景,包括城市与农村地区,以及多种天气和光照条件。

VisDrone2019数据集中包含约540 000个手工标注的目标,涵盖了多个类别,包括行人、人、汽车、面包车、公交车、卡车、摩托车、自行车、遮阳三轮车和三轮车。对于数据集中的每个目标,尺度比常规的目标检测数据集更小,且干扰因素更为复杂,是一个具有较高挑战性的数据集。

因此,本文针对无人机视角下的小目标检测选择该数据集进行实验,用以衡量模型的性能。

3.2 实验环境与评价指标

本文训练和测试所使用的GPU为Nvidia RTX3090,操作系统为Ubuntu,版本为20.04,采用Python3.8编程语言,PyTorch1.8.2深度学习框架,CUDA版本为11.4。训练的学习率为0.01,最大迭代次数为300,早停轮数为15, batch-size为64,优化器为SGD, momentum为0.937,图片输入大小为640×640。

本文使用几个目标检测常用的评估指标来分析改进模型的检测性能,包括精确率(Precision, P)、召回率(Recall, R)、平均精度均值mAP和参数量(Params)。

(1) 精确率(Precision, P)。在所有检测出的边界框中,真正包含目标边界框所占的比例。其计算公式为:

$$P = \frac{TP}{TP + FP} \quad (6)$$

式中:TP是真正例(True Positives)的数量;FP是假正例(False Positives)的数量。

(2) 召回率(Recall, R)。在所有应被检测出的目标中,实际被检测出的边界框所占的比例。其计算公式为:

$$R = \frac{TP}{TP + FN} \quad (7)$$

式中:FN是假负例(False Negatives)的数量。

(3) 平均精度均值(mean Average Precision, mAP)。对于每个类别,都会计算出一个Precision-Recall曲线,然后计算这个曲线下的面积(Area Under the Curve, AUC),得到的值就是该类别的Average Precision(AP)。最终,mAP就是所有类别AP值的平均值。其中,mAP@0.5是所有类别的IoU阈值在0.5时的平均检测精度;mAP@0.5:0.95是以步长为0.05,计算IoU阈值在0.5~0.95之间的所有IoU阈值下的平均检测精度。其计算公式为:

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (8)$$

式中: AP_i 指的是第*i*类检测目标的平均精度;*k*指的是数据集中所有类别的数量。

(4) 参数量(params)。模型中可训练参数的总数,通常包括卷积层的权重以及偏置项等。这些参数的数量直接影响了模型的复杂度和容量,即模型能够学习的信息量。

3.3 实验结果分析

3.3.1 对比实验结果分析

为验证改进算法的网络性能,本文综合评估了多种无人机航拍图像目标检测模型,并通过对比分析深入探讨了它们的性能差异。实验结果如表1所示,详细展示了VisDrone2019数据集中10个类别的mAP@0.5以及总mAP@0.5。实验结果表明,本文的模型检测精度优于大多数当前主流的目标检测模型。与基线模型YOLOv5s相比,本文的模型将总体的mAP@0.5提升了6.6个百分点。小目标类

别包括行人、人、自行车、三轮车和遮阳篷三轮车比基线模型分别高出 10%、6.5%、3.5%、4.1% 和 2.02%；其他目标类别包括汽车、面包车、卡车、巴士和摩托车比基线模型分别高出 8.6%、9%、5.9%、10.1% 和 7%。实验结果表明，本文算法不仅在小目标类别上实现了显著性能提升，在其他目标类别上也取得了不同程度的检测精度提升，总体都优于基线模型 YOLOv5s。与由官方提供的 YOLO 系列网络最

新的 YOLOv8s 相比，本文算法虽然 $mAP@0.5$ 与 YOLOv8s 持平，且在各类汽车目标上精度不如 YOLOv8s，但是这类目标在 visDrone 数据集中属于大目标，而本文主要是针对小目标进行的改进。在主要小目标类别行人、人和自行车中，分别高出 7.6%、6.3% 和 3.1%，与 YOLOv8 相比有明显优势，且更具有针对性。此外，本文改进的算法模型参数量少于 YOLOv8s，在算法部署中也有一定优势。

Table 1 Comparative experimental results on the VisDrone dataset

表 1 在 VisDrone 数据集上的对比实验结果

算法	mAP@0.5/%	行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳篷三轮车	巴士	摩托车
Faster R-CNN	21.7	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.70	31.4	20.7
RetinaNet	13.9	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.20	17.8	11.8
YOLOv5s	32.3	39.8	32.3	10.6	73.1	33.2	26.5	18.4	9.38	41.5	38.0
YOLOv7-tiny	36.1	41.5	37.4	10.0	77.6	37.5	29.6	22.8	11.0	48.8	45.4
YOLOv8s	38.9	42.2	32.5	11.0	79.8	45.4	35.8	27.0	16.0	54.8	44.7
ours	38.9	49.8	38.8	14.1	81.7	42.2	32.4	22.5	11.4	51.6	45.0

3.3.2 消融实验结果分析

为验证各项改进的有效性，本文在 VisDrone2019 数据集上进行了消融实验。以 YOLOv5s 为基线模型，在该模型的基础上逐个增加本文所提出的各项改进。首先，将 CI-OU 损失函数替换为 EIOU 损失函数；然后，在检测头部分加入额外的小目标检测头；最后，引入本文设计的 DC3 模块。如表 2 所示，每一项改进都在不同程度上提高了网络的检测精度。EIOU 损失函数使模型在参数量不变的情况下， $mAP@0.5$ 提升了 1%，召回率提升了 2.2%，说明 EIOU 损失函数能够使模型优化，且其相似性的能力也得到了提升。引入小目标检测头显著提升了模型的检测精度、准确率和召回率，说明该模型有效增强了对数据集中大量小目标的检测能力。DC3 模块使检测精度和召回率都有了 0.4 个百分点的提升，说明骨干网络捕捉重要语义信息的能力也有了一定的提升。综上，本文所改进后的算法相较于基线模型 YOLOv5s 在检测精度、准确率和召回率 3 个方面都有了大幅的提升，尤其是对小目标的检测效果尤为显著。

Table 2 Ablation experiments based on YOLOv5s model

表 2 以 YOLOv5s 为基础模型的消融实验

算法	mAP@0.5/%	mAP@0.5:0.95/%	Precision	Recall	Params/M
YOLOv5s	32.3	17.4	44.3	33.3	7.04
YOLOv5s+EIOU	33.3	17.6	42.6	35.5	7.04
YOLOv5s+EIOU+extra head	38.5	21.0	49.4	38.4	7.18
YOLOv5s+EIOU+extra head+DC3	38.9	21.4	47.2	39.8	9.41

3.3.3 检测结果可视化分析

为了更直观地对比模型改进前后的检测效果，本文从 VisDrone2019 数据集中随机选取了 1 张图片进行检测分析。检测结果如图 a 和图 b 所示，其中图 a 为 YOLOv5s 的检测结果，图 b 为本文改进后模型的检测结果。实验结果表明，在 3 组图片中，左侧图片展示的 YOLOv5s 模型检测结果均存在不同程度的漏检和误检现象，而本文改进模型

有效解决了上述问题，改进后的模型检测效果明显优于改进前的模型。尤其是在图片中距离拍摄点较远的部分，本文模型对小目标的检测明显更加精确，充分验证了本文模型检测效果的优越性。



(a) Detection result of YOLOv5s

(a) YOLOv5s 检测结果

(b) Detection result of the algorithm

(b) 本文算法检测结果

Fig. 3 The detection results of YOLOv5s and the algorithm proposed in this paper

图 3 YOLOv5s 与本文算法检测结果

4 结语

针对当前无人机航拍图像检测效果不佳的问题，本文通过对 YOLOv5s 目标检测模型进行一系列改进，以提升检测精度。通过对主干网络中的 C3 模块进行改进，提出了一个 DC3 模块，用于增强主干网络的特征提取能力。通过将 CI-OU 损失函数优化为 EIOU 损失函数，提升了模型训练

效果。通过增加额外的小目标检测头,提升了对数据集中大量小目标的检测精度。实验结果表明,改进后的模型与基线模型 YOLOv5s 相比, mAP@0.5 提升了 6.6 个百分点, mAP@0.5:0.95 提升了 4 个百分点。本文将在下一步研究中不断优化现有方法,包括提升模型的泛化能力,优化模型的实时性,开展轻量化研究,增强该方法的有效性,以期推动目标检测技术与无人机的进一步结合。

参考文献:

- [1] HONG D, GAO L, YOKOYA N, et al. More diverse means better: multi-modal deep learning meets remote-sensing imagery classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(5): 4340–4354.
- [2] HONKAVAARA E, SAARI H, KAIVOSOJA J, et al. Processing and assessment of spectrometric, stereoscopic imagery collected using a light-weight UAV spectral camera for precision agriculture [J]. Remote Sensing, 2013, 5(10): 5006–5039.
- [3] MENOVAR H, GUVENC I, AKKAYA K, et al. UAV-enabled intelligent transportation systems for the smart city: applications and challenges [J]. IEEE Communications Magazine, 2017, 55(3): 22–28.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580–587.
- [5] 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.
- [6] 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, 2016: 779–788.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C]//14th European Conference on Computer Vision, 2016: 21–37.
- [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980–2988.
- [9] CHEN C, ZHANG Y, LV Q, et al. Rrnet: a hybrid detector for object detection in drone-captured images [C]//Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [10] LIU Z, GAO G, SUN L, et al. HRDNet: high-resolution detection network for small objects [C]//2021 IEEE International Conference on Multimedia and Expo (ICME), 2021: 1–6.
- [11] ZHANG X, IZQUIERDO E, CHANDRAMOULI K. Dense and small object detection in uav vision based on cascade network [C]//Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [12] CHEN J, SUN J, LI Y, et al. Object detection in remote sensing images based on deep transfer learning [J]. Multimedia Tools and Applications, 2022, 81: 12093–12109.
- [13] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778–2788.
- [14] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022: 443–459.
- [15] SU J, QIN Y C, JIA Z, et al. A small object detection algorithm based on ATO-YOLO [J]. Computer Engineering and Applications, 2024, 60(6): 68–77.
苏佳,秦一畅,贾泽,等. 基于ATO-YOLO的小目标检测算法[J]. 计算机工程与应用, 2024, 60(6): 68–77.
- [16] LI G Q, SUN Y J. Target detection in drone aerial photography based on improved SSD [J]. Space Control Technology and Applications, 2023, 49(1): 65–73.
李国强,孙英家. 基于改进SSD的无人机航拍目标检测[J]. 空间控制技术与应用, 2023, 49(1): 65–73.
- [17] WU M J, YUN L J, CHEN Z Q, et al. Improved YOLOv5s small target detection algorithm from the perspective of unmanned aerial vehicles [J]. Computer Engineering and Applications, 2024, 60(2): 191–199.
吴明杰,云利军,陈载清,等. 改进YOLOv5s的无人机视角下小目标检测算法[J]. 计算机工程与应用, 2024, 60(2): 191–199.
- [18] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [DB/OL]. <https://arxiv.org/abs/2004.10934>.
- [19] HUANG G, LIU Z, VAN L, et al. Densely connected convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700–4708.
- [20] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146–157.
- [21] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results [C]//Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

(责任编辑:陈维捷)