DOI:10.19651/j.cnki.emt.2519396

基于 RT-DETR 的无人机航拍图像小目标检测算法

刘 杰 李志文 张腾庆 谢明山
(贵州大学大数据与信息工程学院 贵阳 550025)

摘 要: 随着无人机应用场景不断拓展,航拍图像中小目标检测成为计算机视觉领域的研究热点。针对小目标特征不明显、背景复杂导致误检和漏检,现有算法检测精度与实时性难以兼顾等问题,本研究提出了一种基于 RT-DETR 的航拍图像小目标检测算法 FST-RTDETR 来解决这些问题。首先,将 FasterNet 与 EMA 注意力机制结合,重新设计原有模块的 Basic Block 模块的结构,实现提高网络运行速度和视觉任务的准确性。其次,为了解决传统 P2 检测层添加后出现计算量过大、后处理更加耗时等问题,本研究基于原本的 CCFM 架构上提出使用 P2 特征层经过 SPDConv 得到富含小目标信息的特征给到 P3 进行融合,然后使用 CSP 思想和基于 Omni-Kernel 进行改进得到 CSP-OmniKernel 进行特征整合,有效地学习从全局到局部的特征表现,最终减少漏检率、误检率和提高小目标的检测性能。最后,为了使得模型简化损失函数计算过程、改进回归效率和精度以及拥有更全面的损失考虑,使用 inner-MPDIoU 替换原来的 GIoU。改进后的算法在 VisDrone2019 数据集上的实验表明,FST-RTDETR 模型实现了 49.6%的 mAP@50,相对于原来的 RT-DETR 模型提高了 2.1%。FST-RTDETR 模型显著提升了无人机图像的目标检测性能,提高了模型效率,对比其他算法表现出了良好的性能。

关键词: 无人机检测;RT-DETR;小目标检测;FasterNet-EMA;SPDConv
中图分类号: TP391.4;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.20

Small object detection algorithm in Drone aerial images based on RT-DETR

Liu Jie Li Zhiwen Zhang Tengqing Xie Mingshan
(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: With the continuous expansion of drone application scenarios, small object detection in aerial images has become a research hotspot in the field of computer vision. In view of the problems that small object features are not obvious, complex backgrounds lead to false detection and missed detection, and the existing algorithms are difficult to balance detection accuracy and real-time performance, this paper proposes an aerial image small object detection algorithm FST-RTDETR based on RT-DETR to solve these problems. First, FasterNet is combined with the EMA attention mechanism, and the structure of the Basic Block module of the original module is redesigned to improve the network operation speed and the accuracy of visual tasks. Secondly, in order to solve the problems of excessive calculation and more time-consuming post-processing after adding the traditional P2 detection layer, this study propose to use the P2 feature layer based on the original CCFM architecture to obtain features rich in small object information through SPDConv and give them to P3 for fusion, and then use the CSP idea and Omni-Kernel to improve CSP-OmniKernel for feature integration, effectively learn the feature performance from global to local, and finally reduce the missed detection rate, false detection rate and improve the detection performance of small objects. Finally, in order to simplify the loss function calculation process, improve regression efficiency and accuracy, and have a more comprehensive loss consideration, this study use inner-MPDIoU to replace the original GIoU. Experiments on the improved algorithm on the VisDrone2019 dataset show that the FST-RTDETR model achieves a detection accuracy of 49.6%, which is 2.1% higher than the original RT-DETR model. The FST-RTDETR model significantly improves the object detection performance of drone images, improves model efficiency, and shows good performance compared to other algorithms.

Keywords: Drone detection;RT-DETR;small object detection;FasterNet-EMA;SPDConv

0 引 言

最近几年,无人机技术凭借其灵活机动、成本低廉以及可在复杂环境中作业等显著优势,在多个领域得到了广泛且深入的应用。从农业领域的作物生长监测[1]、病虫害检测[2],到林业中的森林资源调查[3]、火灾预警[4];从交通领

收稿日期:2025-07-18

域的路况实时监控[5]、违章行为识别,再到安防领域的边境巡逻、突发事件应急响应[6]等,无人机航拍技术为实时监测和分析提供了重要的数据支持[7]。在实际应用过程中,无人机航拍图像目标特征不明显、信息量不足,背景环境复杂多变[8],这些因素共同构成了无人机小目标检测的技术难点。

YOLO(you only look once)[9]、SSD(single shot multibox detector)[10]等主流目标检测算法虽能实现高效检测,但在无人机遥感场景中仍存在明显局限。以 YOLO 为例,其检测速度固然出众,却会生成大量冗余框,进而增加后处理的时间成本。近期,基于 Transformer[11]的目标检测算法为航拍图像小目标检测带来了新的思路。RT-DETR(real-time detection transformer)[12]作为其中的典型代表,凭借其高效的检测性能受到广泛关注。

以 YOLO 为代表的检测算法中,虽然具备实时性优势,但在小目标检测过程中会导致小目标特征压缩过度,由于其多尺度特征提取依赖步长为 8/16/32 的下采样,导致小目标细节特征丢失。YOLO 在面临复杂背景时抗干扰能力弱,导致航拍图像中建筑阴影、树木纹理等背景噪声易导致误检。在改进的 YOLO 算法中,Ponduri 等[13]提出了改进的 YOLOv9 模型用于检测极小尺度物体,通过提出 RepNCSPELAN4 模块和 SPPELAN 模块,使 VisDrone 数据集的 mAP@50 达到了 48.7%。Luo 等[14]提出了 EBC-YOLO 模型,通过引入了双向特征金字塔网络来集成特征信息,通过集成 CNF 模块减少下采样和传播过程中的损失,使得 VisDrone 数据集 mAP@50 达到了 44.3%。Li 等[15]提出了一个轻量型的无人机航拍小目标检测算法,通过 C2f 结构与高效注意力机制(efficient attention)相结合,引入了动态头部设计,使得 VisDrone 数据集的 mAP@50 达到了 38.8%,参数量对比 YOLOv8n 减少了 2/3。Qiu 等[16]提出了 YOLO-Air 模型,通过提出 SECAConv(squeeze-excitation convolution with attention)模块来实现动态权重分配和通道注意力机制增强了小物体的特征表示,设计了航空特征金字塔网络来优化特征传输,开发了自适应尺度融合模块来实现提高网络检测小目标的能力,使得 VisDrone 数据集的 mAP@50 达到了 44.5%。翁俊辉等[17]提出了 CS-YOLOv5s 模型,通过引入小目标检测器、最大池化分支嵌入上下文增强模块并注入路径聚合网络,同时采用空间深度转换卷积模块替换下采样卷积模块,使得 VisDrone 数据集 mAP@50 达到了 42.0%。

以 RT-DETR 为代表的检测算法中,虽去除非极大抑制实现端到端检测,但小目标感知能力不足,其默认仅使用 S3/S4/S5,未覆盖高分辨率 P2 层,损失函数对小目标回归不友好,GIoU 损失仅依赖边界框重叠区域优化,对小目标因定位偏差导致的无重叠情况敏感。在改进的 RT-DETR 算法中,Teng 等[18]提出了 Stff-rtdetr,通过引入重新参数化的扩张模块(reparameterized dilation module,DR-Block)、Hilo 注意力机制和跨尺度特征融合金字塔网络(cross-scale feature fusion pyramid network,P2-CCPF),使得 VisDrone 数据集的 mAP@50 达到 39.6%。Han 等[19]提出了 LT-DETR,通过 Cross Stage Partial-Omni Kernel、Dynamic sampling 和 Lightweight Group Convolution Channel Shuffle 3 个模块来优化特征融合,使模型计算量降低了近 50%,VisDrone 数据集的 mAP@50 达到了 37.0%。Liu 等[20]提出了 ESO-DETR 模型,通过加入门控单头注意力主干块(gated single-head attention backbone block,GSHA)来增强局部细节提取,利用多尺度多头自注意力机制(multiscale multihead self-attention mechanism,MMSA)来管理骨干网络中的复杂特征,通过引入高效的特征融合金字塔网络(efficient feature fusion pyramid network,ESO-FPN)网络来增强小目标检测,使得 VisDrone 数据集 mAP@50 达到了 41.0%。刘亚蒙等[21]提出将 RT-DETR 原始 ResNet18 主干提取网络中的 Basic Block 替换为轻量级 FasterNetBlock,加入无参注意力模块(simple attention module,SimAM),使得 mAP@50 提升 3.1%,Params 和 FLOPs 相比于原始的算法分别降低了 15.6%和 13%。张靖雯等[22]提出引入轻量化骨干网络 RE-FasterNet,在小目标检测头中嵌入注意力尺度序列融合框架,使得 mAP@50 提升 2.8%,模型大小和 FLOPs 相比于原始的算法分别降低了 23.6%和 13.1%。

然而,针对航拍图像小目标检测的特殊需求,RT-DETR 仍存在算法检测精度与实时性难以兼顾,小目标检测不精准和面向遮挡物体漏检、俯瞰图出现检测错误等问题。为克服这些存在的问题,进一步提升航拍图像小目标检测的性能,本研究提出了一种基于 RT-DETR 的航拍图像小目标检测算法 FST-RTDETR。本研究算法主要从 3 个方面进行创新改进:

1)提出 FasterNet-EMA 模块,将 FasterNet[23]与高效多尺度注意力(efficient multi-scale attention,EMA)[24]机制相结合,对 Basic Block 模块的结构进行重新设计,在有效提高网络运行速度的同时,显著提升了视觉任务的准确性。

2)针对传统 P2 检测层添加后带来的计算量过大、后处理耗时等问题,基于原本的跨尺度特征融合模块(cross-scale feature fusion module,CCFM)框架进行改进,通过使用 P2 特征层经过空间深度转换卷积[25](space-to-depth convolution,SPDConv)得到富含小目标信息的特征,并与 P3 进行融合,再利用 CSP(Cross Stage Partial)思想和基于 Omni-Kernel[26]改进得到 CSP-OmniKernel 进行特征整合,从而有效学习从全局到局部的特征表现,极大地提高了小目标的检测性能。

3)为简化模型损失函数计算过程,改进回归效率和精度,并实现更全面的损失考虑,使用 MPDIoU[27]结合 Inner-IoU[28]思想替换原来的 GIoU[29]。

通过在 VisDrone2019 数据集上对 FST-RTDETR 模

型进行实验，结果表明，该模型实现了 49.6%的检测精度（mAP@50），相较于原来的 RT-DETR 模型提高了 2.1%。提升无人机图像目标检测性能的同时，提高了模型效率，在与其他算法的对比中展现出良好的性能优势。后续将详细阐述算法的具体设计、实验设置及结果分析。

1　无人机航拍图像小目标检测算法

1.1　RT-DETR

RT-DETR 是由百度飞桨提出的实时端到端目标检测模型，基于 DETR（detection transformer）[30] 框架改进，旨在解决传统 DETR 推理速度慢、计算冗余的问题，同时保留其无需非极大值抑制[31] 的端到端优势。其核心设计包括 3 部分：主干网络（backbone）、混合编码器（hybrid encoder）和解码器（decoder）。

RT-DETR 采用多尺度特征提取主干网络，主干网络输出 3 个尺度的特征图（S3、S4、S5），分别对应输入图像下采样 8 倍、16 倍和 32 倍，与主流检测模型的多尺度设计兼容。混合编码器由注意力增强的尺度内特征交互模块（AIFI）和跨尺度特征融合模块组成，是 RT-DETR 速度优化的核心。解码器沿用 DETR 的 Transformer 结构，但引入 IoU 感知查询选择、自适应推理速度两关键改进。选择的基础模型 RT-DETR 的结构如图 1 所示。

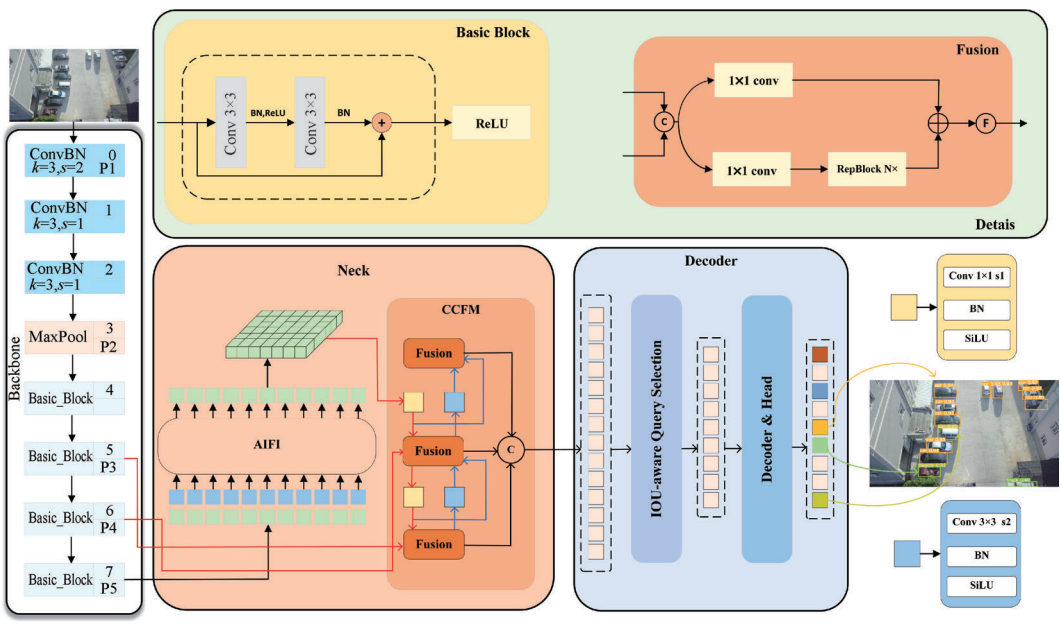


图 1　RT-DETR 模型结构

Fig. 1　RT-DETR network structure

1.2　基于 RT-DETR 模型的改进

本研究所提出的 FST-RTDETR 算法进行了多维度的创新优化。将 FasterNet 与 EMA 注意力机制结合对 Basic Block 模块的改造，是加速与提效的关键。FasterNet 以其轻量化结构减少计算量，EMA 注意力机制则能自适应聚焦目标区域，二者协同优化模块内部的卷积、激活与残差连接流程，使得网络在处理航拍图像时，可快速提取有效特征，在减少参数数量的同时，保障视觉任务的精度。针对 P2 检测层的优化，通过 SPDConv 对 P2 特征层处理，充分挖掘小目标细节信息，将富含小目标特征的数据输送至 P3 层融合，有效避免传统方法的计算冗余。在此基础上，CSP-OKM（csp-omnikernel）整合模块借助 CSP 结构减少重复计算，结合 Omni-Kernel 动态核的多尺度特征捕捉能力，实现对航拍图像从全局场景到局部小目标特征的高效整合，显著增强模型对小目标的感知能力。在损失函数优化上，inner-MPDIoU 优化预测框与真实框的 4 个角点距离，统一考虑重叠区域、中心点距离、宽高偏差。相较于 GIoU，inner-MPDIoU 多部件分解和内部特征点监督，解决了 GIoU 对目标内部结构不敏感的问题，尤其提升了对小目标的检测精度。在航拍图像小目标定位时，能更精准度量二者位置关系，简化回归计算流程，进而提升检测精度与效率，使得 FST-RTDETR 在复杂航拍场景下的小目标检测表现更优。改进后的 FST-RTDETR 网络结构如图 2 所示。

1.3　FasterNet 模块

FasterNet 是 CVPR 2023 提出的高效神经网络架构，旨在解决传统轻量级网络中低 FLOPs 与高延迟的矛盾。其核心创新在于部分卷积（PConv）算子，通过仅对输入通道子集执行常规卷积，大幅降低计算冗余与内存访问，同时保持特征提取能力。

网络架构上，FasterNet 采用分层设计，包含 4 个阶段，每个阶段通过嵌入层或合并层进行下采样与通道扩展，并

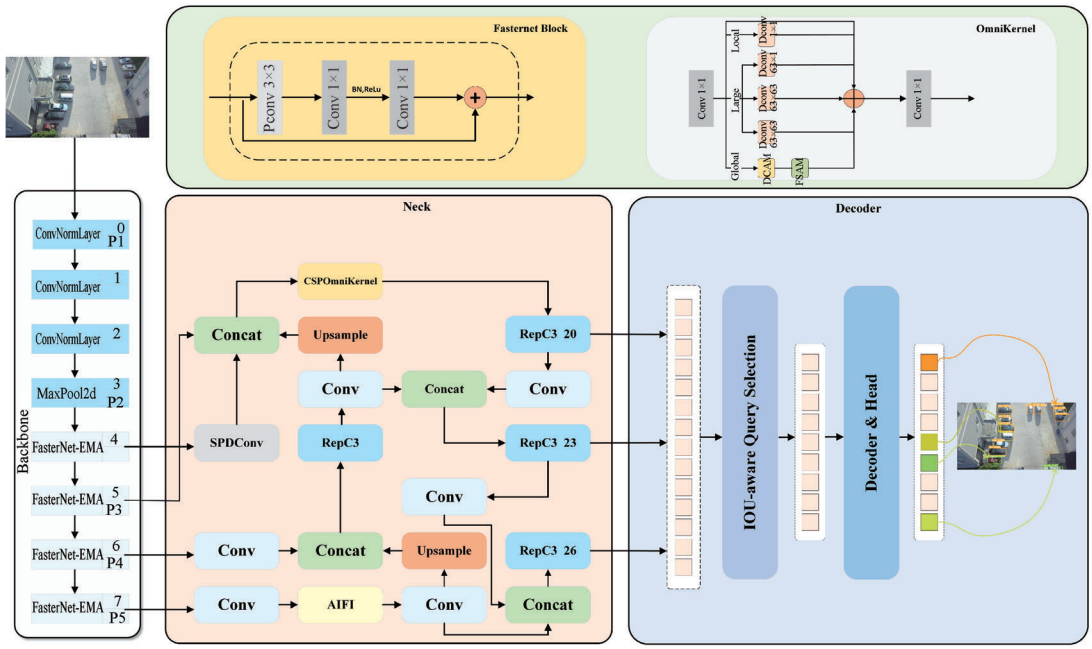


图 2　FST-RTDETR 模型结构

Fig. 2　FST-RTDETR network structure

堆叠多个 FasterNet 块。块内基于 PConv 构建倒残差结构,配合残差连接增强网络稳定性。通过调整网络深度与宽度,FasterNet 提供多种变体,适配不同计算资源需求。FasterNet Block 的模块结构如图 3 所示。

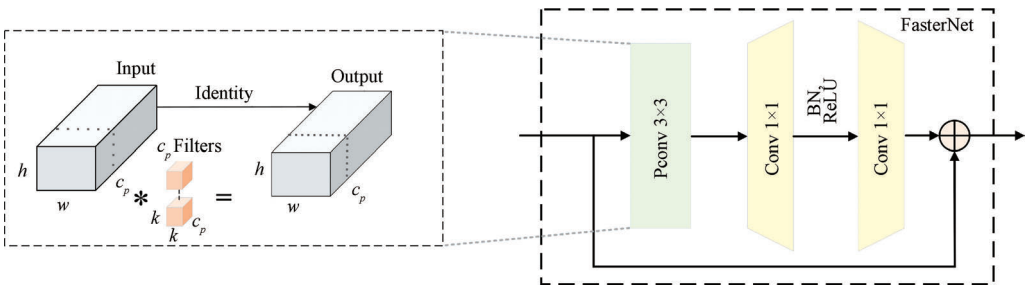


图 3　FasterNet 模块结构

Fig. 3　FasterNet module structure

1.4　跨空间学习的高效多尺度注意力模块

跨空间学习的高效多尺度注意力模块是一种创新性注意力机制模块。它致力于高效提取深度视觉特征,规避传统通道降维导致的信息损耗。该模块将部分通道重塑至批量维度,并对通道维度分组,促使空间语义特征均匀分布于各特征组。

模块内设置两个并行分支,EMA 采用了协调注意力(coordinate attention,CA)[32]模块中的共享 1×1 卷积分支,并将其命名为 1×1 分支。此外,为了聚合多尺度的空间结构信息,EMA 在 1×1 分支并行放置了一个 3×3 卷积核,命名为 3×3 分支。考虑到特征分组和多尺度结构,EMA 能够有效地建立短程和长程依赖,从而提高性能。在图像分类、目标检测等视觉任务中,该模块能够显著提升模型对多尺度目标的感知与定位能力,以较低计算开销增强模型性能。EMA 的模块结构如图 4 所示。

在航拍图像小目标检测任务中,EMA 注意力通过积累多帧或多尺度特征的统计特性,增强小目标的特征表示,同时抑制复杂背景干扰,提升检测精度,体现出高效性与鲁棒性的优势。

1.5　FasterNet EMA 模块

为了提升航拍图像小目标检测任务的多尺度特征融合能力以及解决传统轻量级网络中低 FLOPs 与高延迟的矛盾,将 FasterNet 和 EMA 模块进行融合。FasterNet-EMA 的结构如图 5 所示。

1.6　小目标检测层

在目标检测任务中,小目标在常规特征金字塔网络(FPN)的 P3、P4、P5 层上常面临表征能力不足的挑战。传统解决方案通过引入 P2 检测层增强小目标特征,但会显

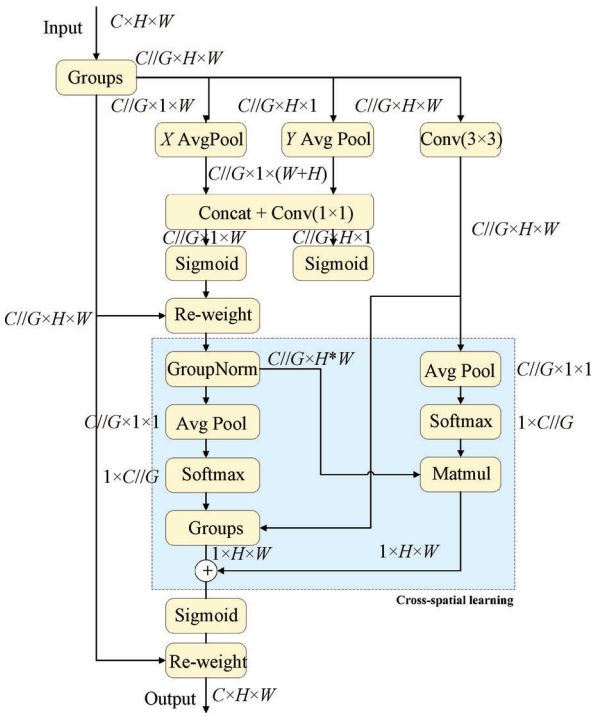


图 4　EMA 模块结构

Fig. 4　EMA module structure

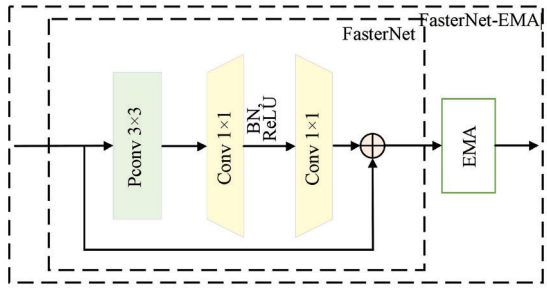


图 5　FasterNet-EMA 模块结构

Fig. 5　FasterNet-EMA module structure

著增加计算复杂度与后处理时延。为平衡检测性能与效率，本研究基于 CCFM 框架提出改进方案：首先采用 SPDConv 处理 P2 层特征，提取高分辨率小目标信息并融合至 P3 层以强化细节表征，SPDConv 工作原理如图 6(c)所示；随后结合 CSP[33] 思想与 Omni-Kernel 结构构建特征整合模块，CSP-OKM 结构如图 6(b)所示。该模块通过 3 个并行分支—全局分支、大尺度分支、局部分支实现从全局到局部的多尺度特征协同学习，在避免 P2 层直接引入计算负担的同时，显著提升小目标检测精度。Omni-Kernel 结构如图 6(a)所示。最后构建的小目标检测层如图 7 所示。

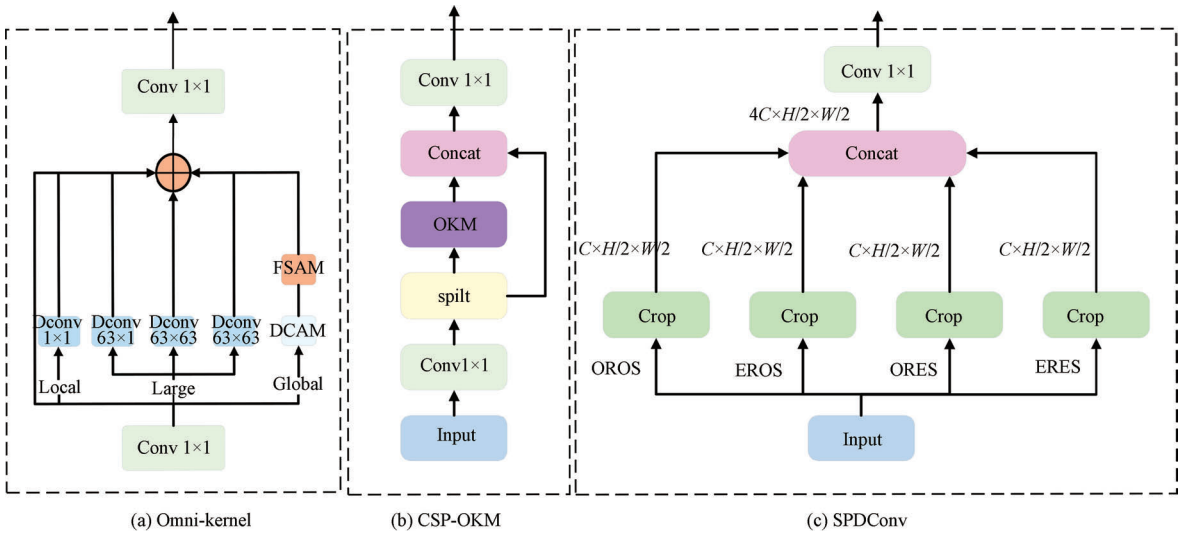


图 6　SPDConv、Omni-Kernel 及 CSP-OKM 结构

Fig. 6　SPDConv, Omni-Kernel and CSP-OKM structure

1.7　损失函数

RT-DETR 采用的边界框回归损失函数是 GIoU Loss，如式(1)～(2)所示。其中 A、B 是目标框和预测框的面积，C 是两个框的最小外接矩形的面积。

$$GIoU = IoU - \frac{|C-(A \cap B)|}{C} \tag{1}$$

$$L_{GIoU} = 1 - GIoU \tag{2}$$

Inner-MPDIoU 是对 MPDIoU 损失的改进，通过引入辅助边界框和尺度因子(Ratio)解决传统 IoU 损失在高、低质量样本上的回归效率问题。其中，尺度因子 ratio∈[0.5,1.5]动态控制辅助框的缩放比例。给定真实框与预测框的中心坐标与宽高参数后，相关计算公式如式(3)～(9)所示。

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2} \tag{3}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \tag{4}$$

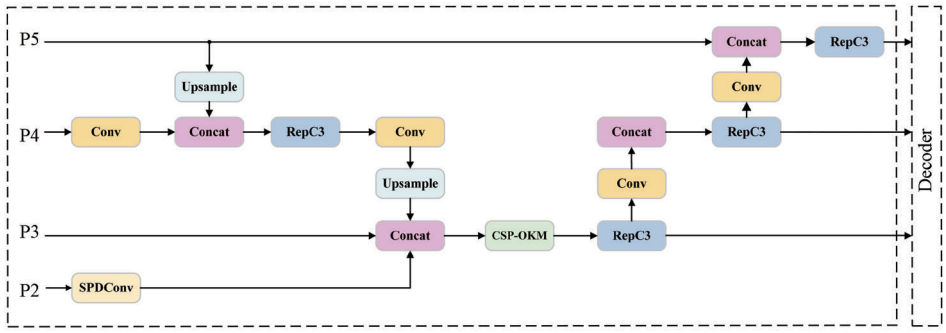


图 7　小目标检测层

Fig. 7　Small object detection layer

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \tag{5}$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \tag{6}$$

$$intersection = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \tag{7}$$

$$Union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - intersection \tag{8}$$

$$IoU^{inner} = \frac{intersection}{union} \tag{9}$$

ratio∈[0.5,1]时,即为使用较小尺度的辅助边框计算 IoU 损失,这将有助于高 IoU 样本回归,达到加速收敛的效果。Inner-IoU 示意图如图 8 所示。

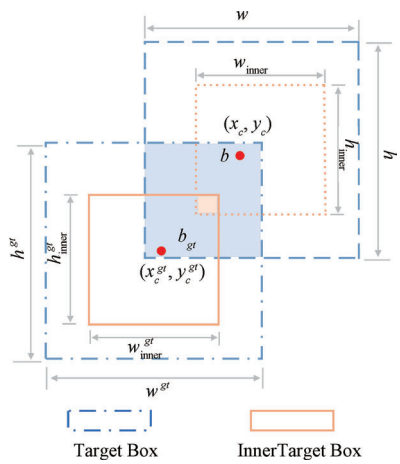


图 8　ratio<1 时 Inner-IoU 示意图

Fig. 8　Inner-IoU diagram when ratio<1

ratio∈[1,1.5]时,即为使用较大尺度的辅助边框计算 IoU 损失,能够加速低 IoU 样本回归过程。Inner-IoU 示意图如图 9 所示。

ρ 代表计算两点间的欧氏距离。b,b^{gt} 分别代表预测框和真实框的中心点坐标,c 表示能够同时包含预测框和真实框的最小边界框的对角线长度。结合 Inner 思想,Inner-DIoU 损失函数可以定义为式(10):

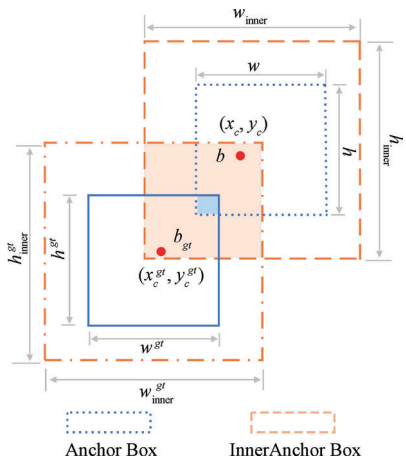


图 9　ratio>1 时 Inner-IoU 示意图

Fig. 9　Inner-IoU diagram when ratio>1

$$L_{Inner-DIoU} = 1 - IoU^{inner} + \frac{\rho^2(b, b^{gt})}{C^2} \tag{10}$$

为了使得模型简化损失函数计算过程、改进回归效率和精度以及拥有更全面的损失考虑,使用 inner-MPDIoU 替换原来的 GIoU。

2　实　　验

2.1　实验数据集和实验环境

本研究实验数据使用由天津大学机器学习和数据挖掘实验室 AISKYEYE 团队收集的大规模无人机视觉数据集 VisDrone2019[34]数据集。这些数据集来自于各类无人机摄像头,采集于中国 14 个不同城市,相隔数千公里。覆盖了城市和农村等不同环境,包含行人、车辆、自行车等不同物体,以及稀疏和拥挤等不同密度场景,是在不同天气和光照条件下,使用不同型号的无人机平台收集的。超过 260 万个感兴趣的目标框被手工标注,标注对象包括行人、汽车、自行车和三轮车等。同时还提供了场景可见性、对象类别和遮挡情况等重要属性,以帮助更好地利用数据。静态图像中有 6 471 张用于训练,548 张用于验证,1 610 张用于测试。

网络实验环境基于 Windows 10，GPU 选用 RTX 4070tisuper，显存为 16 G，选用 PyTorch2.3.1，训练时间为 200 个 epoch，选用 Python 3.10.16，选用 Cuda 11.8，输入图像尺寸为 640×640。训练过程中，初始学习率为 0.000 1。实验环境配置如表 1 所示。

表 1　实验环境和配置信息

Table 1　Experimental environment and configuration information

配置类型	配置名称	配置信息
软件配置	操作系统	Windows
	Python 版本	3.10.16
	Pytorch	2.3.1
	CUDA	11.8
	Cudnn	8.9.6
硬件配置	CPU	AMD Ryzen 7900X
	GPU	NVIDIA 4070ti super
	显存大小	16 GB

2.2　评价指标

在固定实验条件下，通过对比模型增强前后的图像检测效果差异来评估算法性能。采用精确率(Precision)、召回率(Recall)、F1 分数、平均精确率(mAP)和 GFLOPs 作为评估标准。

精度表示的是预测为正的样本中有多少是真正的正样本，预测结果中真正的正例的比例。精度如式(11)所示。

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

召回率表示的是样本中的正例有多少被预测正确了，所有正例中被正确预测出来的比例。召回率如式(12)所示。

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

F1 分数是准确率和召回率之间的调和平均值，目的是将这些指标整合到一个指标当中，避免 Precision 或 Recall 的单一极大值，用于综合反映整体的指标。F1 分数在 0～1 变化，当分数越接近 1 的时候就代表模型更好。F1 分数的计算如式(13)所示。

$$F1 = \frac{(2 \times Precision \times Recall)}{Precision + Recall} \tag{13}$$

AP 是单个类别的精确率-召回率(PR)曲线下面积，反映模型对某一类别的检测能力，公式如式(14)所示。mAP 是将各分类的 AP 取平均值，如式(15)所示。

$$AP = \int_0^1 P(r)\mathrm{d}r \tag{14}$$

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{15}$$

GFLOPs(giga floating-point operations per second)是衡量目标检测模型计算复杂度的核心指标，表示模型每秒执行的十亿次浮点运算量，GFLOPs 用于量化模型对硬件算力的需求，直接影响推理速度和部署可行性。

2.3　小目标检测层计算量实验

为了科学评估小目标检测层对减少计算量的具体表现，基于 VisDrone 2019 数据集，针对基线模型开展系统的计算量实验。实验结果汇总于表 2。

表 2　计算量汇总

Table 2　Summary of calculation quantity

类别	计算量/GFLOPs	参数量/10^6
RT-DETR-r18	57.0	19.9
RT-DETR-r18＋P2	81.7	18.9
RTDETR＋小目标检测层	58.6	20.1
FST-RTDETR	59.7	17.5

相较于传统 P2 检测层，首先对 P2 特征层进行 SPDConv 处理，从中提取富含小目标信息的特征并输送至 P3 层完成特征融合，引入 CSP 核心思想并结合 Omni-Kernel 进行改进，构建出 CSP-OKM 模块用于特征整合。该方案能够有效学习从全局到局部的完整特征表征。由表 2 可以看出在计算量上，小目标检测层相对于传统 P2 检测层减少了 23.1 GFLOPs。计算量仅在基准模型上增加了 1.6 GFLOPs，参数量仅在基准模型上增加了 0.2×10^6。最终的模型在计算量和参数量上相较于添加普通的 P2 检测层有明显的下降。

2.4　消融实验

为科学评估改进算法的有效性，基于 VisDrone 2019 数据集，针对基线模型开展系统的消融实验，实验结果汇总于表 3。

实验结果表明，相较于基线模型，改进后的 FST-RTDETR 模型(模型 E)在目标检测精度上实现显著提升，mAP@50 指标提高 2.1%。具体来看，在模块替换实验中，实验 A 将原始基础模块替换为 FasterNet-EMA 模块，mAP@0.5 指标提升 0.6%，参数量下降了 3×10^6，浮点数运算下降了 5.5 GFLOPs，验证了该模块在加速网络运行的同时，能够有效增强视觉任务处理的准确性。在框架改进实验 B 中，通过优化 CCFM 框架，mAP@50 指标提升 1.3%，这一结果表明改进后的框架能够更高效地学习从全局到局部的特征表达，进而显著提升小目标检测性能。在损失函数改进实验 C 中，mAP@50 指标提升 0.6%。在实验 D 中，结合 FasterNet-EMA 模块和小目标检测层后，mAP@50 指标提升 1.5%，参数量减少了 2.4×10^6。最终构建的模型 E，在参数量减少 2.4×10^6，浮点数运算仅增加 2.7 GFLOPs 的情况下，实现 mAP@50 指标 2.1%的提

表 3　消融实验结果

Table 3　Ablation experiment results

实验	FasterNet-EMA	小目标检测层	Inner-MPDIoU	mAP@50/%	参数量/10^6	GFLOPs
RT-DETR-r18				47.5	19.9	57.0
A	√			48.1	16.9	51.5
B		√		48.8	20.5	65.2
C			√	48.1	19.9	57.0
D	√	√		49.0	17.5	59.7
E	√	√	√	49.6	17.5	59.7

升。并且检测速度达到 37.0 fps,满足了实时性检测的标准。这一结果充分证明,改进后的 FST-RTDETR 模型在未显著增加计算复杂度的前提下,能够有效提升目标检测精度,展现出良好的算法优化效果。

2.5　各个类别检测精度对比

为量化评估 FST-RTDETR 算法的检测性能提升效果,在 VisDrone2019 数据集上,对基准算法 RT-DETR 与改进算法 FST-RTDETR 进行了多类别检测精度的系统性对比分析,具体实验结果如表 4 所示。该表完整呈现了 Pedestrian、People、Bicycle、Car、Van、Truck、Tricycle、Awning-tricycle、Bus、Motor 共 10 个目标类别的 mAP@50 指标,以及综合 mAP@50 统计值。实验数据表明,FST-RTDETR 算法在所有 10 个检测类别中,mAP@50 指标均优于原 RT-DETR 算法,充分验证了该算法对不同尺度目标检测任务的良好适应性。上述实验结果有力支撑了 FST-RTDETR 算法在实际目标检测任务中的优越性,为相关领域的技术应用提供了可靠的性能依据。

表 4　各个类别检测精度对比

Table 4　Comparison of detection accuracy of each category

类别	mAP@50/%		
	RT-DETR	FST-RTDETR	increase
pedestrian	55.8	**57.5**	1.7
people	48.3	**51.0**	2.7
bicycle	21.0	**21.5**	0.5
car	85.6	**86.4**	0.8
van	50.5	**51.6**	1.1
truck	38.8	**40.6**	1.8
tricycle	33.7	**37.5**	3.8
awning-tricycle	18.6	**21.4**	2.8
bus	62.7	**67.5**	4.8
Motor	59.9	**61.0**	1.1
all	47.5	**49.6**	2.1

2.6　基础模型对比实验

为客观验证 FST-RTDETR 算法在小目标检测领域的性能优势,本研究以 VisDrone2019 数据集为测试基准,选取双阶段 R-CNN 系列与单阶段 YOLO 系列中的代表性先进算法作为基线模型开展对比实验,具体结果如表 5 所示。

表 5　各种模型实验对比

Table 5　Comparison of various model experiments

类别	mAP@50/%						
	Faster R-CNN	YOLOv5	YOLOv6	YOLOv8	YOLOv11	RT-DETR	FST-RTDETR
pedestrian	21.4	35.0	29.7	36.2	36.9	55.8	**57.5**
people	15.6	27.5	24.6	28.9	28.6	48.3	**51.0**
bicycle	6.7	8.41	4.29	8.9	9.97	21.0	**21.5**
car	51.7	75.3	73.9	76.3	76.4	85.6	**86.4**
van	29.7	38.7	36.3	40.4	37.7	50.5	**51.6**
truck	19.0	31.3	24.7	32.2	30.5	38.8	**40.6**
tricycle	13.1	22.4	18.0	24.1	23.1	33.7	**37.5**
awning-tricycle	7.7	11.7	11.1	13.5	13.3	18.6	**21.4**
bus	31.4	49.1	41.9	49.4	49.7	62.7	**67.5**
Motor	20.7	35.8	30.9	38.2	38.2	59.9	**61.0**
all	21.7	33.5	29.7	34.8	34.4	47.5	**49.6**

实验针对数据集涵盖的 10 个目标类别进行系统性检测性能评估,数据显示 FST-RTDETR 算法在各目标类别检测中均展现出显著性能优势。特别在小目标检测任务中,面对纹理特征相似、形态多变的 Bicycle 和 Awning-tricycle 类别,该算法仍能实现 21.5% 和 21.4% 的 mAP@50 检测精度,充分体现其对复杂小目标的特征提取与识别

能力。在大尺寸目标检测方面，FST-RTDETR 算法同样保持良好性能平衡，相比原始 RT-DETR 模型，在 Car 和 Truck 类别检测精度上分别提升 0.8%和 1.8%。最终，FST-RTDETR 算法在综合 mAP@50 指标上达到 49.6%，相较于基线模型实现显著提升。上述实验结果表明，通过改进特征提取网络结构与多尺度融合机制，FST-RTDETR 算法有效增强了对小目标的特征捕捉与定位能力，在无人机航拍图像复杂场景下的目标检测任务中展现出显著性能优势，为实际工程应用提供了可靠的技术支撑。

2.7　改进模型对比实验

为客观验证 FST-RTDETR 算法在小目标检测领域的性能优势，本研究以 VisDrone2019 数据集为测试基准，选取各种改进的 YOLO 模型和改进的 RT-DETR 模型展开对比实验，结果如表 6 所示。

表 6　改进模型实验对比

Table 6　Comparison of improved model experiments

类别	mAP@50/%				
	Efficient YOLOv9	EBC-YOLO	Stff-RTDETR	ESO-DETR	FST-RTDETR
pedestrian	53.2	51.2	41.0	42.4	**57.5**
people	46.6	41.1	29.3	31.1	**51.0**
bicycle	35.5	17.8	16.2	17.0	21.5
car	76.6	83.8	78.8	78.8	**86.4**
van	52.4	47.9	38.9	40.1	51.6
truck	47.0	41.8	46.7	47.8	40.6
tricycle	40.6	31.4	24.9	26.5	**37.5**
awning-tricycle	26.7	16.8	19.3	21.7	**21.4**
bus	61.1	59.1	58.4	59.8	**67.5**
Motor	47.5	52.1	18.4	44.9	**61.0**
all	48.7	44.3	39.6	41.0	**49.6**

实验围绕不同目标检测模型展开，对 pedestrian、people、bicycle 等多类目标及所有目标的检测性能进行对比。FST-RTDETR 在多数目标类别上表现出色。在 pedestrian 类别中，其 mAP@50 高达 57.5%，远高于 Efficient YOLOv9 的 53.2%、EBC-YOLO 的 51.2%等其他模型。people 类别里，FST-RTDETR 以 51.0%的 mAP@50 领先其他模型。car 类别中，FST-RTDETR 更是取得 86.4%的检测精度，在该类别检测上优势突出。bus 类别下，67.5%的成绩也远超其他模型。在综合的 mAP@50 下以 49.6%高于其他模型。相较之下，Stff-RTDETR 虽在 truck 类别表现较为优秀，但是整体表现欠佳，多个类别 mAP@50 处于较低水平。FST-RTDETR 在此次改进模型实验的目标检测性能上更具优势，检测效果更优。

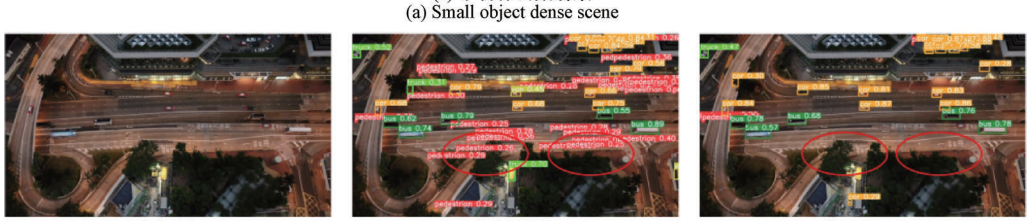
2.8　可视化检测结果对比

为客观评估提出的 FST-RTDETR 算法在无人机航拍场景下的检测性能，从 VisDrone2019 数据集中筛选出具有代表性的复杂场景数据，包括小目标密集分布、高空俯瞰视角以及夜间航拍 3 类典型场景，开展与基准算法 RT-DETR 的对比实验。其中，小目标聚集和高空俯瞰视角具有挑战性的场景。实验结果如图 10 所示，图像组从左至右依次为原始航拍图像、RT-DETR 模型检测结果、FST-RTDETR 算法检测结果，关键差异区域通过圆框线进行标注。

Original image　　RT-DETR　　FST-RTDETR

(a) 小物体密集场景

(a) Small object dense scene



Original image　　RT-DETR　　FST-RTDETR

(b) 高空鸟瞰图

(b) Aerial view from high altitude

图 10　各类场景下目标检测效果对比

Fig. 10　Object detection effect comparison in various scenarios

在小目标密集分布场景下,量化分析表明,FST-RTDETR 算法的漏检率显著低于基准算法。具体而言,该算法成功检测出基准算法未能识别的 8 辆遮阳三轮车和 14 辆摩托车,展现出对小目标更强的检测能力。针对高空俯瞰视角场景,FST-RTDETR 算法有效降低了错检率,如图 10 中圆圈区域所示,RT-DETR 模型将地面补丁误判为行人,而改进算法未出现此类误检情况。在夜间航拍场景中,FST-RTDETR 算法表现出对多尺度目标变化的良好适应性,能够准确检测不同尺寸的物体,相较于原始模型,其错检概率明显降低。

综上所述,通过在 VisDrone2019 数据集典型复杂场景下的对比实验,FST-RTDETR 算法在目标检测性能上全面优于基准算法 RT-DETR,尤其在高空俯瞰视角场景下展现出更强的鲁棒性,验证了算法改进策略的有效性与实用性。

利用热图评估了改进模型和基准模型的性能,评估结果如图 11 所示。可以观察到以下几点:在处理密集的人群场景时,热图中小物体的响应非常强烈和清晰,几乎每个目标都被清楚地识别出来。对比在夜间场景和光照充足的场景,夜间场景虽然背景光照较差,但热力图中虚线框内的车辆等小物体仍然被检测出,准确率相对较高。光照充足场景对比基准模型,准确率也相对较高,错检率相对较低。

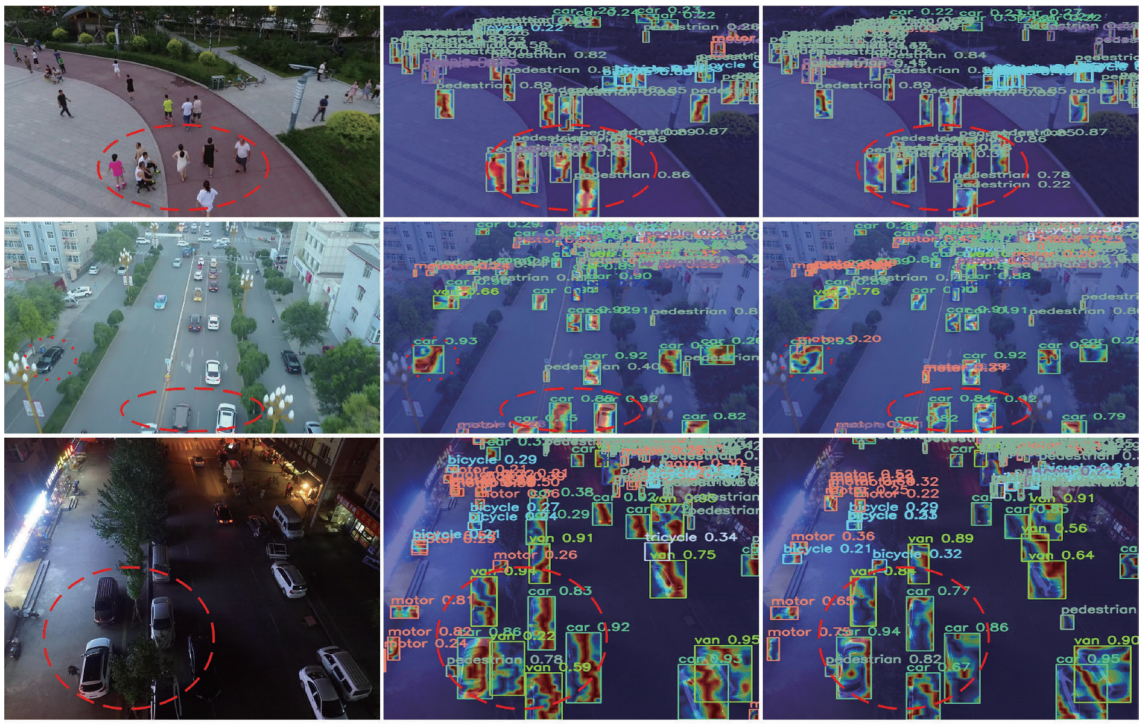


图 11　热力图测试结果比较

Fig. 11　Comparison of thermal map test results

3　结　　论

针对无人机航拍图像中普遍存在的小目标特征不明显、背景干扰复杂所导致的模型误检率高和漏检率高问题，以及现有算法在检测精度与实时性之间难以有效权衡的挑战，从特征提取与融合策略优化、小目标检测层结构设计等维度对 RT-DETR 模型进行了改进。所提方法在提升算法检测性能的同时，有效降低了模型的参数量。

为应对小目标检测的核心挑战，重构了 RT-DETR 的骨干网络（Backbone），引入了 EMA 注意力机制及 FasterNet 模块，以提升网络的运行效率和视觉特征提取能力。针对传统 P2 检测层引入后导致计算量剧增、后处理耗时显著上升的问题，在原有 CCFM 架构基础上，提出改进策略：将 P2 特征层经 SPDConv 处理，得到富含小目标语义信息的特征后与 P3 层融合；并引入 CSP 结构思想和基于 Omni-Kernel 的改进机制，设计出 CSP-OmniKernel 模块进行特征整合。该模块能有效学习从全局到局部的多层次特征表征，最终提升小目标检测性能。最后，对损失函数进行了优化改进，在简化计算过程的同时，提升了边界框回归的精度和效率，并增强了损失函数的全面性考量。

在 VisDrone2019 数据集上的实验评估表明，改进后的算法在检测精度上取得了显著提升，兼具参数更少、精度更高的优势，基本满足实时应用需求。然而，算法在特定复杂场景下仍存在漏检和误检现象，未来研究将聚焦于进一步提升网络的鲁棒性，并将本算法扩展应用于更复杂多变的实际场景。

参考文献

[1] HUANG Y N, QIAN Y R, WEI H Y, et al. A survey of deep learning-based object detection methods in crop counting[J]. Computers and Electronics in Agriculture, 2023, 215: 108425.

[2] WANG SH H, XU D CH, LIANG H J, et al. Advances in deep learning applications for plant disease and pest detection: A review[J]. Remote Sensing, 2025, 17(4): 698.

[3] BERIE H T, BURUD I. Application of unmanned aerial vehicles in earth resources monitoring: Focus on evaluating potentials for forest monitoring in Ethiopia[J]. European Journal of Remote Sensing, 2018, 51(1): 326-335.

[4] MANOJ S, VALLIYAMMAI C. Drone network for early warning of forest fire and dynamic fire quenching plan generation[J]. EURASIP Journal on Wireless Communications and Networking, 2023, 2023(1): 112.

[5] KHAN N A, JHANJHI N Z, BROHI S N, et al. Smart traffic monitoring system using unmanned aerial vehicles (UAVs)[J]. Computer Communications, 2020, 157: 434-443.

[6] WANKMULLER C, KUNOVJANEK M, MAYRGUNDTER S. Drones in emergency response-evidence from cross-border, multi-disciplinary usability tests[J]. International Journal of Disaster Risk Reduction, 2021, 65: 102567.

[7] WU W N, LIU AO, HU J W, et al. EUAVDet: An efficient and lightweight object detector for UAV aerial images with an edge-based computing platform[J]. Drones, 2024, 8(6): 261.

[8] ZHANG Q, ZHANG H Y, LU X W. Adaptive feature fusion for small object detection[J]. Applied Sciences, 2022, 12(22): 11854.

[9] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZALEZ J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.

[11] HAN K, XIAO AN, WU EN H, et al. Transformer in transformer[J]. Advances in Neural Information Processing systems, 2021, 34: 15908-15919.

[12] ZHAO Y AN, LYU W Y, XU SH L, et al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[13] PONDURI V, MOHAN L, BASHEERA S, et al. Highly efficient YOLOv9 model for detecting extremely small-scale objects [J]. Engineering Research Express, 2025, 7(1): 015287.

[14] LUO H K, WANG Y Q, CHEN Y L, et al. EBC-YOLO: A remote sensing target recognition model adapted for complex environments[J]. Earth Science Informatics, 2025, 18(3): 282.

[15] LI X L, BAO Y F. Small target detection algorithm for UAV aerial photography based on improved YOLO V8n[C]. 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS). IEEE, 2024: 870-875.

[16] QIU J G, CAI F K, FU N, et al. YOLO-Air: An efficient deep learning network for small object detection in drone-based imagery[J]. IEEE Access, 2025, 13: 79718-79735.

[17] 翁俊辉，成乐，黄曼莉，等. 基于 CS-YOLOv5s 的无人机航拍图像小目标检测[J]. 电子测量技术，2024，

47(7):157-162.
WENG J H, CHENG L, HUANG M L, et al. Small target detection in drone aerial images based on CS-YOLOv5s[J]. Electronic Measurement Technology, 2024, 47(7):157-162.

[18] TENG X X, ZHANG W D, LIU T, et al. Stff-rtdetr: A small object detection algorithm based on drone aerial photography[J]. The Journal of Supercomputing, 2025, 81(8): 928.

[19] HAN Z X, JIA D L, ZHANG L. LT-DETR: Lightweight UAV object detection and dual knowledge distillation for remote sensing scenarios[J]. Measurement Science and Technology, 2025, 36(3): 036005.

[20] LIU Y F, HE M, HUI B. ESO-DETR: An improved real-time detection transformer model for enhanced small object detection in UAV imagery[J]. Drones, 2025, 9(2): 143.

[21] 刘亚蒙,赵友全,孙振涛,等.构建改进 RT-DETR 算法检测隐形眼镜环状波纹缺陷[J]. 电子测量与仪器学报,2024,38(5):1-9.
LIU Y M, ZHAO Y Q, SUN ZH T, et al. Constructing an improved RT-DETR algorithm for detecting circular ripple defects in contact lenses[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5): 1-9.

[22] 张靖雯,孙坚,徐红伟,等.基于改进 RT-DETR 的玻璃绝缘子缺陷检测算法[J]. 电子测量技术,2025, 48(14):96-105.
ZHANG J W, SUN J, XU H W, et al. Glass insulator defect detection algorithm based on improved RT-DETR[J]. Electronic Measurement Technology, 2025, 48(14): 96-105.

[23] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.

[24] OUYANG D L, HE S, ZHANG G ZH, et al. Efficient multi-scale attention module with cross-spatial learning[C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2023: 1-5.

[25] 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. Cham: Springer Nature Switzerland, 2022: 443-459.

[26] CUI Y N, REN W Q, KNOLL A. Omni-kernel network for image restoration[C]. AAAI Conference on Artificial Intelligence, 2024, 38(2): 1426-1434.

[27] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv:2307.07662, 2023.

[28] ZHANG H, XU C, ZHANG SH J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[J]. ArXiv preprint arXiv: 2311.02877, 2023.

[29] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.

[30] ZHU X ZH, SU W J, LU L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection[J]. ArXiv preprint arXiv:2010.04159, 2020.

[31] HOSANG J, BENENSON R, SCHIELE B. Learning non-maximum suppression[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4507-4515.

[32] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[33] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.

[34] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019: 213-226.

作者简介

刘杰,硕士研究生,主要研究方向为深度学习、图像处理,目标检测。
E-mail:17785100921@163.com

李志文,硕士研究生,主要研究方向为深度学习、目标检测,空间域知识图谱。
E-mail:79614027@qq.com

张腾庆,硕士研究生,主要研究方向为深度学习、目标检测。
E-mail:1484315072@qq.com

谢明山(通信作者),教授,博士生导师,主要研究方向为机器人、空间域知识图谱、深度学习。
E-mail:mingshanxie317@163.com