

DOI:10.19651/j.cnki.emt.2520107

改进 RT-DETR 的无人机小目标检测算法^{*}符强^{1,2} 彭章伟^{1,2} 纪元法^{1,2,3,4} 任风华^{1,2}

(1. 桂林电子科技大学广西精密导航技术与应用重点实验室 桂林 541004; 2. 桂林电子科技大学信息与通信学院 桂林 541004;
3. 时空信息与智能位置服务国际合作联合实验室 桂林 541004; 4. 广西产研院时空信息技术研究所有限公司 南宁 530023)

摘要: 针对无人机视角下目标尺度小、遮挡严重导致的特征提取困难及检测精度偏低问题,提出了一种改进 RT-DETR 的无人机小目标检测算法 SwiftHawk-DETR。首先,重构特征金字塔网络(FPN)的融合路径,新增微小目标检测层 P2 以强化细粒度特征捕捉,移除冗余大目标检测层 P5 以精简计算;其次,设计可变形快速多尺度注意网络 Dfaster_net,通过动态采样机制强化小目标边界细节捕捉,同时采用高效特征提取策略降低计算开销;然后,在颈部网络中引入小波特征升级(WFU)缓解原模型上采样与 Concat 操作导致的边缘高频特征失真,结合 SlimNeck 架构优化跨尺度特征融合;最后,构建 Focaler EIou 与 NWD 加权融合的损失函数,通过平衡定位误差与特征分布差异提升小目标建模能力。实验结果表明,算法在 VisDrone2019 和 HIT-UAV 数据集上,mAP50 和 mAP50:95 较基准 RT-DETR 分别提升 3.8% 和 4.2%,参数量降低 61%。

关键词: RT-DETR;无人机目标检测;跨尺度特征融合;动态采样;小波特征升级

中图分类号: TP391.41;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 510.6040

Improved UAV small object detection algorithm based on RT-DETR

Fu Qiang^{1,2} Peng Zhangwei^{1,2} Ji Yuanfa^{1,2,3,4} Ren Fenghua^{1,2}

(1. Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin 541004, China;
2. School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China;
3. Joint Laboratory of International Cooperation on Spatio-Temporal Information and Intelligent Location Services, Guilin 541004, China;
4. Guangxi Institute of Industry and Research Spatio-Temporal Information Technology Co., Ltd., Nanning 530023, China)

Abstract: To address the problems of difficult feature extraction and low detection accuracy caused by small target scales and severe occlusion from the UAV perspective, this paper proposes an improved RT-DETR-based UAV small target detection algorithm named SwiftHawk-DETR. Firstly, the fusion path of the feature pyramid network (FPN) is reconstructed; A tiny target detection layer P2 is added to enhance fine-grained feature capture, and the redundant large target detection layer P5 is removed to simplify computations. Secondly, a deformable fast multi-scale attention network (Dfaster_net) is designed, which strengthens the capture of small target boundary details through a dynamic sampling mechanism while adopting an efficient feature extraction strategy to reduce computational overhead. Thirdly, wavelet feature upgrade (WFU) is introduced into the neck network to alleviate the distortion of edge high-frequency features caused by upsampling and Concat operations in the original model and the slimneck architecture is combined to optimize cross-scale feature fusion. Finally, a loss function based on the weighted fusion of Focaler EIou and NWD is constructed to improve the small target modeling ability by balancing localization errors and feature distribution differences. Experimental results on the VisDrone2019 and HIT-UAV datasets show that compared with the baseline RT-DETR, the proposed algorithm increases mAP50 and mAP50:95 by 3.8% and 4.2% respectively, and reduces the number of parameters by 61%.

Keywords: RT-DETR; UAV small target detection; cross-scale feature fusion; dynamic sampling; wavelet feature upgrade

0 引言

随着计算机视觉与无人机技术的迅猛发展,基于无人

机的航拍目标检测技术日趋成熟,广泛应用于灾害救援、农业监测、电力巡检、城市安防等多个领域。然而,无人机航拍场景的复杂性为目标检测带来了诸多挑战:高空作业导

收稿日期:2025-10-25

^{*} 基金项目:广西科学技术厅项目(AB23026120)、南宁市科学研究与技术开发项目(20231029)、广西科技计划项目(桂科 AA23062038,桂科 ZY23055048,桂科 AB23026120,桂科 AA24206025,桂科 AA24206043)、国家自然科学基金(U23A20280,62161007,62471153)、产研计划项目(CYY-HT2023-JSJJ-0023-1,CYY-HT2023-JSJJ-0024-1)、广西科技基地和人才专项(桂科 AD25069103)资助

致目标尺寸小、像素占比低,特征提取困难;复杂背景干扰、目标遮挡重叠及高密度分布等问题进一步加剧了检测难度。传统目标检测方法如基于方向梯度直方图(histogram of oriented gradients, HOG)、尺度不变特征变换(scale-invariant feature transform, SIFT)的手工特征框架在此场景下表现受限,不仅因人工设计特征难以适配小目标弱特征表达而精度低、复杂背景适应性差,还受限于无人机的算力资源,模型轻量化程度无法满足实时检测需求。

目前,深度学习驱动的目标检测器主要分为两大类。一是以卷积神经网络(convolutional neural network, CNN)为基础的模型,经历了从两阶段到单阶段的演进。两阶段方法(如 R-CNN、Faster R-CNN)通过生成候选框再识别目标,精度较高但速度较慢;单阶段方法如 YOLO(you only look once)、单次多框检测器(single shot multibox detector, SSD)直接预测结果,速度快但精度稍逊。YOLO 系列因其在速度与精度上的平衡,端到端训练及低背景误检率,在工业界广受青睐。针对无人机航拍场景,研究者提出了多项改进:赵爽等^[1]通过大尺寸图像切分和多尺度特征融合提升小目标检测能力;张伟等^[2]采用层间通道级联实现高效特征复用,改进空间金字塔池化(spatial pyramid pooling, SPP)以增强小目标细节提取;Zhang 等^[3]提出上下文双向增强(contextual bidirectional enhancement, CBD-E)策略,抑制背景干扰并强化目标特征;张瑞芳等^[4]构建双向加权多尺度动态颈部网络,促进深浅层特征融合并新增 P2 检测头。这些改进提升了单阶段模型在航拍场景的性能,但仍依赖锚框设计、阈值筛选及非极大值抑制(non-maximum suppression, NMS),导致正负样本不平衡、超参数敏感及鲁棒性受限,难以实现真正的端到端检测。二是以 Transformer 为基础的检测模型,为解决 CNN 局限性提供了新思路。2020 年 Carion 等^[5]提出检测变压器(detection transformer, DETR),利用 Transformer 的全局特征提取能力,通过可学习对象查询和二分局匹配实现无 NMS 的端到端检测,但其参数量大且训练时间较长。Zhu 等^[6]提出 Deformable DETR,引入可变形注意力有效降低了计算成本。

2023 年百度飞桨团队提出实时端到端目标检测器 RT-DETR^[7-8],设计了高效混合编码器和交并比(intersection over union, IoU)感知查询,实现实时 Transformer 目标检测,性能超越 YOLO 系列等主流模型。为了进一步提高小目标检测性能,张晞等^[9]提出多尺度边界特征增强协同网络 MBFECN 以改进颈部网络,通过多尺度边界特征增强机制和高效特征融合策略,提升模型对小目标的感知能力以及整体检测性能,但该方法的多尺度池化、多次卷积及边界增强模块的特征处理引入了额外计算开销,且对大目标检测性能有轻微牺牲。向毅伟等^[10]结合维度感知选择性整合模块(dimension-aware selective integration module, DASI)及尺度序列特征融合模块(scale

sequence feature fusion module, SSFF)提出 DSSF 特征融合结构,充分利用特征图之间的相关性,丰富特征融合阶段的小目标信息,减少关键信息丢失,提高小目标检测精度,但 3D 堆叠卷积与自适应特征选择机制引入了额外的参数量和计算量,导致模型 FPS 下降,实时性受损;损失函数部分提出将 Inner-IoU 的思想融入广义交并比(generalized intersection over union, GIoU)中,引入辅助边框计算损失,加速收敛,但未突破 IoU 类损失在极端遮挡场景下的固有局限性。程鑫淼等^[11]提出了一种差异增强网络(enhanced difference net, EDNet)改进主干,通过门控机制动态调整特征权重,实现“特征博弈”,提升模型对前景目标的敏感性,抑制背景干扰,但该方法在处理极端复杂遮挡场景时存在局限,且面对小规模数据集的泛化性不足。

为了综合解决现有方法在计算效率、模型泛化性、复杂场景适应性、小目标定位精度与鲁棒性方面的不足。本研究将 RT-DETR-R18 作为基准模型,提出改进 RT-DETR 的无人机目标小检测算法 SwiftHawk-DETR,主要改进为:

1)重构特征金字塔融合路径并设计高效特征提取网络。针对微小目标尺度特性,新增高分辨率检测层 P2 以强化细粒度特征捕捉,移除冗余的大目标检测层 P5 以精简计算;

2)设计可变形快速多尺度注意力网络(Dfaster_net),引入可变形卷积网络(deformable convolutional networks v3, DCNv3),通过其动态采样机制精准捕捉小目标的边界细节。同时,设计高效多尺度注意力(Efficient multi-scale attention, EMA)机制结合 FasterNet 架构优化特征提取策略,有效减少计算冗余和模型参数量。

3)提出窄颈小波特征融合网络(slimneck wavelet feature fusion network, SWFM),通过小波特征升级(wavelet feature upgrade, WFU)技术,有效缓解上采样和 Concat 操作引起的边缘高频特征失真问题。同时,引入 slimneck 架构优化特征融合,有效提升跨尺度特征交互效率,减少计算冗余。

4)设计 Focaler EIoU 与归一化高斯距离(normalized wasserstein distance, NWD)加权融合的损失函数,实现几何约束与分布相似性的互补,进一步提升模型在复杂遮挡环境中的小目标定位能力与抗干扰性能。

1 RT-DETR 模型

RT-DETR 是首个实时端到端目标检测模型,在性能上实现双重突破:一是在速度与精度的平衡上超越主流实时检测器,兼顾高精度与实时性;二是无需后处理(如非极大值抑制),从根本上避免后处理延迟,确保推理速度稳定。该模型包含 6 个版本,本文选取网络体量大、运行效率高且适配小目标检测场景的 RT-DETR-R18 作为基准模型,网络结构如图 1 所示^[12]。

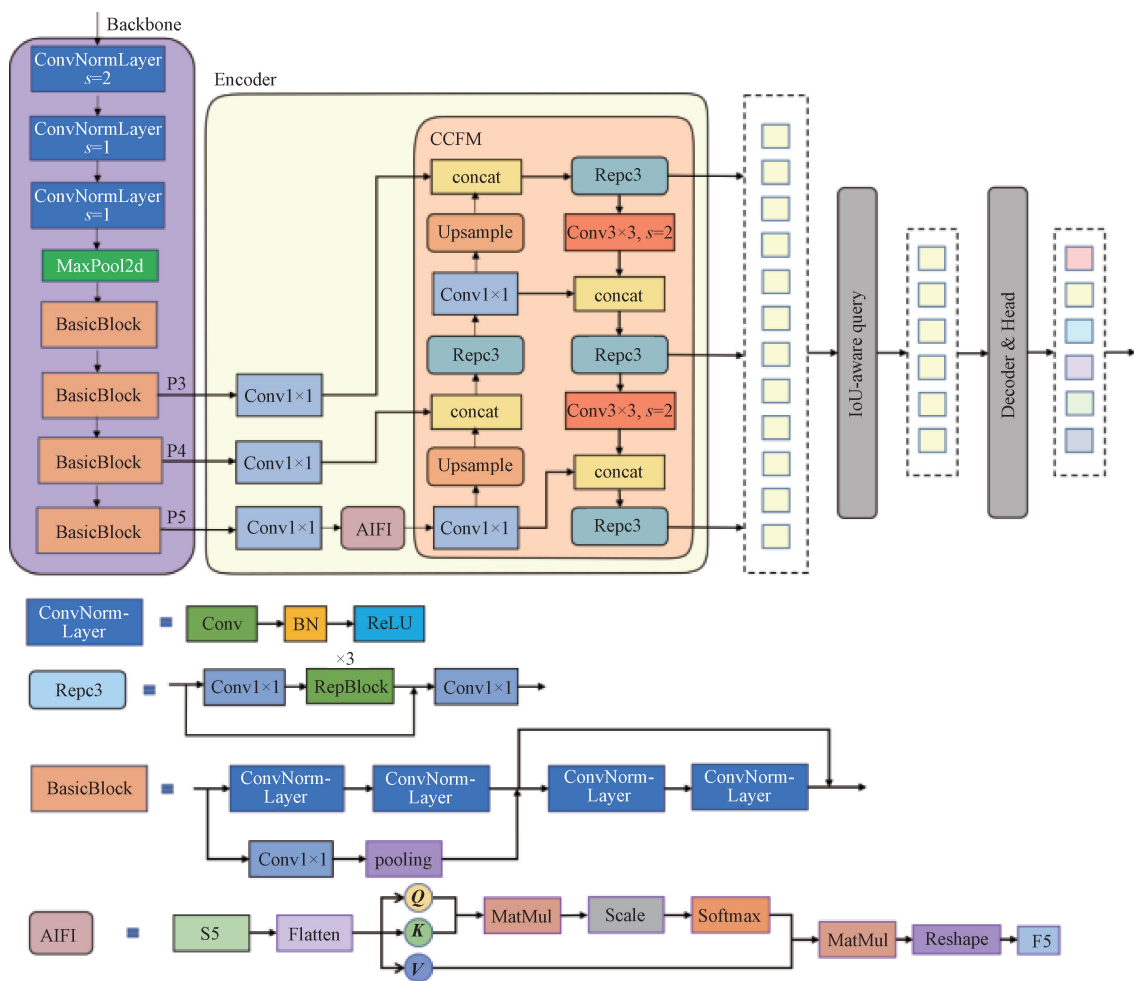


图 1 RT-DETR-R18 网络结构

Fig. 1 RT-DETR-R18 network structure

从架构层面来看,RT-DETR 基于 Transformer 设计,有效解决了多尺度特征处理的难题。其整体结构由 3 部分组成:首先,ResNet-18 主干网络通过卷积层与 BasicBlock 完成初始特征的提取;其次,混合编码器融合了基于注意力的尺度内特征交互 (attention-based intra-scale feature interaction, AIFI) 与跨尺度特征融合模块 (cross-scale feature fusion module, CCFM),通过双向融合机制及上、下采样操作,高效地整合多尺度信息;最后,解码器包含 IoU 感知查询模块,用于筛选高置信度的预测框,从而提升检测精度。此外,解码器与 Decoder head 通过迭代优化,生成高精度的边界框及置信度分数,最终实现高效的检测效果^[7]。

2 SwiftHawk-DETR

针对无人机航拍检测任务中存在的背景复杂、小目标样本多、遮挡严重等问题,本文选用 RT-DETR 作为基线模型,从特征融合路径、主干网络、颈部网络、损失函数 4 个方面对 RT-DETR 进行改进,提出了 SwiftHawk-DETR 模

型,其结构如图 2 所示。

2.1 特征融合路径重构

基于 640×640 的默认输入图像尺寸,以像素面积为基准,无人机航拍目标的尺度标准定义为微小目标 ($< 16 \times 16$)、小目标 ($16 \times 16 \sim 32 \times 32$)、中目标 ($32 \times 32 \sim 64 \times 64$) 及大目标 ($> 64 \times 64$)。VisDrone 数据集的目标分类与尺寸分布如图 3 所示。图 3(a) 显示小目标 (如行人和汽车) 占比极高接近 90%,而大目标 (如卡车、公交车) 占比不足 10%;图 3(b) 进一步表明中、小、微小目标为数据主体。且即使是数据集中最大的目标 (如卡车),其尺度也多在 64×64 以内,归类为中目标。

基准算法 RT-DETR 的主干网络输出 20×20 、 40×40 和 80×80 这 3 个尺度的特征图,分别对应于大、中、小目标的检测。传统改进方法为增强小目标检测能力,通常通过增设 P2 层 (160×160) 并与深层特征进行融合 (方法 1)^[4],虽可提升检测性能,但会显著增加计算开销。

结合无人机数据集目标以中、小和微小目标为主,且大目标尺寸大多在 64×64 pixels 以内的特点,P4 层已足

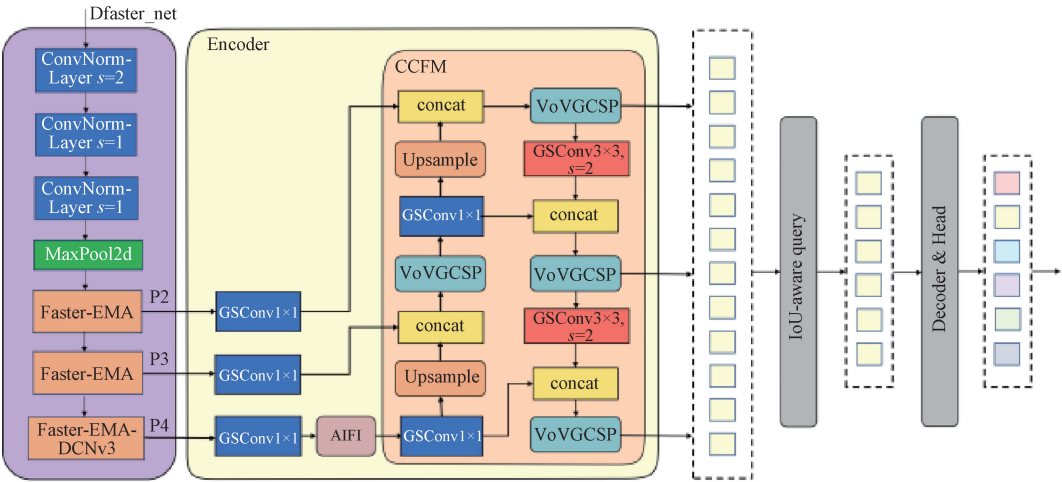
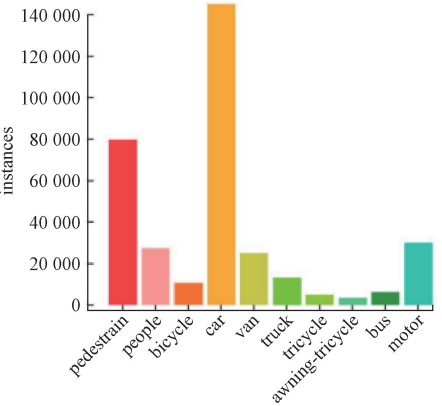
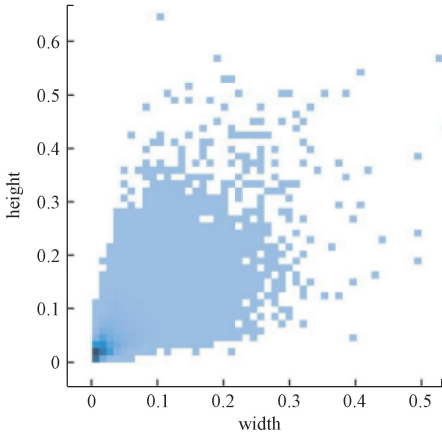


图 2 SwiftHawk-DETR 模型结构
Fig. 2 Structure of the SwiftHawk-DETR model



(a) 各类别实例数量柱状图
(a) Bar chart of instance counts by category



(b) 目标宽度与高度的散点密度
(b) Scatter plot of target width and height

图 3 目标分类情况和尺寸分布

Fig. 3 Object category distribution and size distribution

的提升作用有限,导致资源利用率较低。

因此,本研究提出重构特征融合路径:在增设微小目标检测层 P2 的同时,移除大目标检测层 P5,并在特征融合模块中移除与之相关的融合操作(方法 2)。

两种方法的对比结果如表 1 所示。实验表明,与方法 1 相比,方法 2 在检测精度 mAP50 仅降低 0.2% 的前提下,参数量和千兆浮点计算量(giga floating-point operations per second, GFLOPs)分别减少了 55% 和 11.3%。方法 2 在极大减少参数量和计算量的情况下,仍能保持较高的检测精度,实现了更高效的特征提取,验证了方法的可行性^[13]。

表 1 两种改进方法对比

Table 1 Comparison of two improvement methods						
算法	P/ %	R/ %	mAP50/ %	mAP50:95/ %	Prma/ ($\times 10^6$)	GFLOPs
RT-DETR	62.6	46.7	48.3	29.7	19.9	57.0
方法 1	62.8	48.7	50.0	31.0	20.0	81.0
方法 2	62.9	48.2	49.8	30.6	8.9	71.8

2.2 Dfaster_net 主干网络

由于 RT-DETR 参数量大且存在大量计算冗余,为实现推理速度与检测精度的协同优化,本文构建了 Dfaster_net 主干网络。通过设计快速多尺度特征提取网络,增加特征提取的效率,将原 Basicblock 替换为效率更高的 Faster-EMA 模块,并在深层引入 DCNv3 可变形卷积增强,扩大感受野,增强小目标定位能力。结构设计如下。

1) 主体结构设计

该模块将轻量级网络架构 FasterNet 所采用的高效计算单元 FasterBlock,与 EMA 进行融合。通过对特征提取的路径进行重构,在降低整体计算复杂度的同时,进一步

强化了对多尺度特征的融合效能。FasterBlock 的核心在于引入部分卷积(partial convolution, PConv)算子,其采用通道选择性计算策略:仅对输入特征图的局部连续通道子集(首/末个通道)执行空间卷积,保留其余通道恒等映射。这种通道维度的稀疏计算机制以极低代价捕获全局空间信息,在维持表征能力的前提下实现计算资源与内存带宽的协同优化^[14-15]。图 4 为 Faster-EMA 模块和 Pconv 与普通卷积的对比。

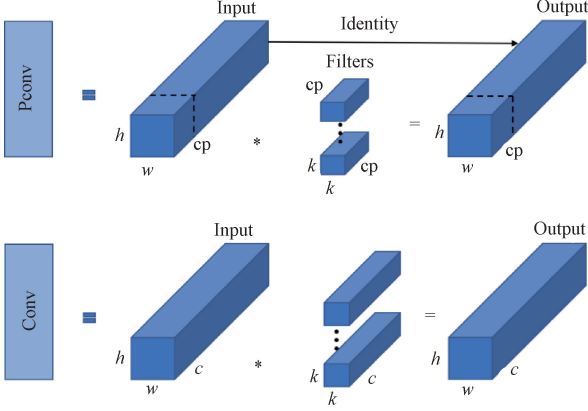


图 4 Faster-EMA 块和 Pconv 卷积与普通卷积的对比

Fig. 4 Faster-EMA block and comparison of Pconv convolution and standard convolution

因此 Pconv 的 FLOPS 和内存访问量为:

$$h \times w \times k^2 \times c_p^2 \quad (1)$$

$$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \quad (2)$$

普通卷积的 FLOPS 和内存访问量为:

$$h \times w \times k^2 \times c^2 \quad (3)$$

$$h \times w \times 2c + k^2 \times c^2 \approx h \times w \times 2c \quad (4)$$

式中: (c, h, w) 为输入通道尺寸, h 和 w 分别代表输入特征图的高度和宽度, k 为卷积核大小、 c 为输入通道数, c_p 为 Pconv 的输入通道数^[16]。一般默认 $c_p = \frac{1}{4}c$ 。

由上式可以得出 Pconv 的 FLOPS 为普通卷积的 $\frac{1}{16}$,

内存访问量为普通卷积的 $\frac{1}{4}$ 。

此外,为了进一步提高对小目标的检测性能,本文引入 EMA 注意力机制。其核心在于跨空间学习策略,通过整合通道信息生成高分辨率注意力权重,显著增强小目标关键特征捕捉。EMA 的计算流程如图 5 所示,其核心由双 1×1 卷积分支与 3×3 卷积分支构成多路径特征处理架构。双 1×1 分支首先执行全局平均池化操作,其数学表达为:

$$\text{AvgPool} = \frac{1}{H \times W} \sum_j \sum_i x_c(i, j) \quad (5)$$

其中, H 和 W 分别为输入特征图的高度与宽度, $x_c(i, j)$ 表示第 c 通道的像素值。

池化结果经 1×1 卷积降维后,采用 Sigmoid 函数进行非线性变换,随后通过重加权运算(Re-weight)实现通道间的特征融合,促进 1×1 分支之间的信息交互。 3×3 分支利用 3×3 卷积提取多尺度上下文信息,与 1×1 分支输出形成互补特征。 1×1 分支经重加权后的结果通过组归一化(groupnorm, GN)处理,随后与 3×3 分支的输出分别经过全局平均池化和 Softmax 函数完成特征归一化。归一化后的特征与原始分支输出执行点积运算,生成保留空间位置信息的注意力图。通过跨空间学习模块整合双分支特征,增强全局空间编码能力,最终经 Sigmoid 函数和 Re-weight 运算捕获全局上下文,输出与输入同尺寸的注意力特征图,可无缝集成至各类网络架构。

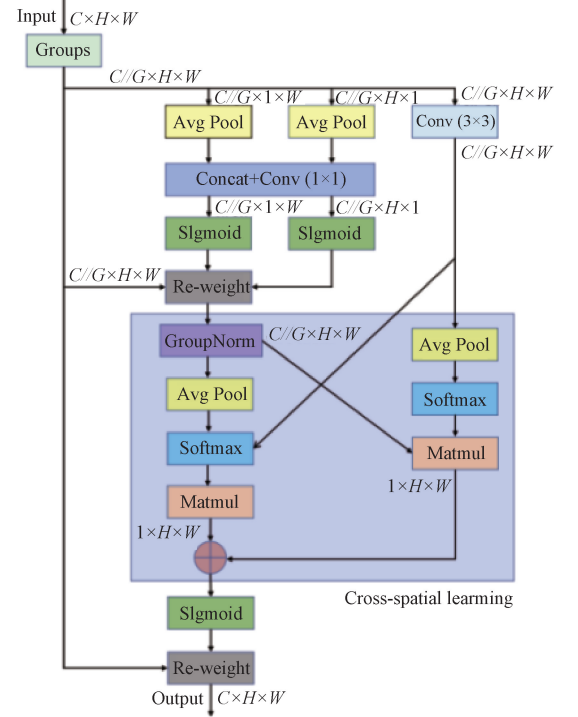


图 5 EMA 计算流程

Fig. 5 Flowchart of EMA calculation

2) 末端变形感知特化

针对检测头对几何变形敏感的特性,在网络最后一层引入 DCNv3,构建 Faster-EMA-DCN 子模块,如图 6 所示。DCNv3 通过动态稀疏采样自适应调整感受野以匹配目标形态,结合多组聚合机制融合语义与空间细节,增强小目标定位鲁棒性与形变目标特征捕捉能力。但其计算复杂度为常规卷积的 2.3 倍,过度使用易致冗余。因此选择语义丰富的 P4 层引入该模块:既借 P4 层丰富的上下文信息与 DCNv3 动态性互补强化小目标定位,又通过感受野扩展弥补去除 P5 层的大目标特征覆盖不足。

DCNv3 核心原理是在训练过程中额外学习采样点的偏移量。通过自适应调整卷积操作的采样位置,DCNv3 能够有效提取图像中复杂的几何特征和多样化的模式,相较

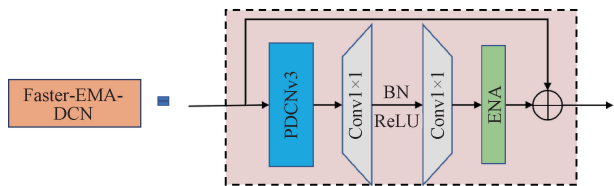


图 6 Faster-EMA-DCN 模块结构

Fig. 6 Faster-EMA-DCN module structure

于传统卷积网络,在处理形状不规则或姿态变化显著的目标时表现出更优的性能。从数学角度看,对于输入特征图 X ,标准卷积在位置 p_0 的输出 $Y(p_0)$ 公式如下:

$$Y(p_0) = \sum_{p_n \in R} W(p_n) \cdot X(p_0 + p_n) \quad (6)$$

其中, $W(p_n)$ 为卷积操作对应的权重参数,其表征了不同采样位置对输出特征的贡献程度; R 定义了卷积核在处理输入特征图时所覆盖的局部邻域范围,确定了参与卷积计算的采样点集合; $X(p_0 + p_n)$ 则表示输入特征图在 $p_0 + p_n$ 位置处的特征值,该值与对应权重 $W(p_n)$ 相乘后累加,形成标准卷积操作的输出结果。

可变形卷积在传统卷积操作的基础上,通过额外学习的偏移量 Δp_n 实现卷积核采样位置的自适应调整。具体而言,可变形卷积在位置 p_0 处的输出特征值 $Y(p_0)$ 由式(7)定义。

$$Y(p_0) = \sum_{p_n \in R} W(p_n) \cdot X(p_0 + p_n + \Delta p_n) \quad (7)$$

其中,偏移量 Δp_n 由额外的卷积层通过学习获得的偏移量 Δp_n ,对每个采样位置 p_n 进行自适应调节,以更好地适应输入图像中目标对象的几何变形特性。通过这种分层增强方式,可变形卷积通过动态调整采样位置,能够高效提取图像中的复杂特征,避免冗余计算,从而显著提升对不规则形状目标的特征捕捉能力^[17]。

2.3 窄颈小波特征融合网络(SWFM)

RT-DETR 的颈部网络 CCFM 虽能实现跨分辨率特征融合,但其“多分支卷积+密集连接”的设计导致参数量和计算量过高,且易引入冗余特征,削弱了关键语义信息的传递效率。此外,其多分辨率特征的交互依赖于残差连接与 Concat 堆叠,前者仅能实现线性特征传递,无法动态适配不同分辨率特征的语义优先级;后者未能解决特征分布差异问题,直接堆叠易导致语义稀释,最终影响小目标与复杂背景下的检测性能。

针对上述问题,本文从两方面对 CCFM 进行改进: 1)采用轻量级卷积方法轻量级卷积方法(global shuffle convolution, GSConv)替换普通卷积,并用 VoVNet-全局混洗卷积空间金字塔池化模块(vovnet-global shuffle convolutional spatial pyramid pooling, VoVGSCSP)替换 Repc3; 2)引入 WFU,优化多分辨率特征的交互机制,实现更精准的特征融合。

1) GSConv 卷积

GSConv^[18]卷积是一种创新的轻量级卷积模块,能够有效解决传统卷积在原始网络中预测计算速度缓慢的问题。它在保持较低时间复杂度的同时,尽可能平衡模型的识别精度与速度。GSConv 卷积的网络结构如图 7 所示,主要由标准卷积、DWConv 深度可分离卷积、Concat 拼接模块和 Shuffle 混洗模块 4 个部分构成。

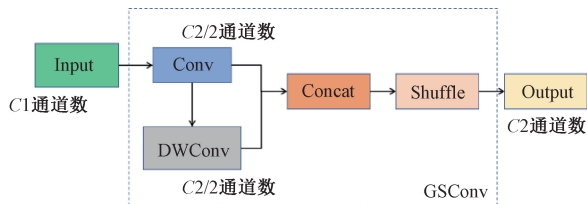


图 7 GSConv 卷积网络结构

Fig. 7 GSConv convolutional network architecture

首先,输入的特征图通过标准卷积进行下采样和特征提取。接着,利用深度可分离卷积(depthwise convolution, DWConv)对每个通道的特征信息进行分别处理。随后,通过 Concat 模块将两种不同卷积提取到的特征在通道上进行拼接。最后,结合 Shuffle 混洗模块,将标准卷积生成的信息与深度可分离卷积的信息进行通道上的交互,从而得到最终的输出特征图。这种改进不仅确保了有效特征提取,还保持了模型的轻量化,极大地提升了整体网络模型的运行效率。

2) VoV-GSCSP 模块

VOV-GSCSP 模块采用一次性聚合方法设计,是一种跨级网络结构,能够在极大程度上减少原模型的计算量和降低网络结构的复杂度,同时依然保持充分的特征提取能力,确保模型具备足够的精度。VOV-GSCSP 模块的网络结构如图 8 所示。该模块的核心构成是通过 GSConv 卷积构建的 GS Bottleneck 模块,如图 8(a)所示。其将上下层的特征图进行拼接后再执行卷积操作,形成了残差块结构,有效避免了网络中的梯度消失问题。此外,多分支卷积的设计增强了特征提取能力和扩展了感受野,进一步提升了模型的识别准确率。

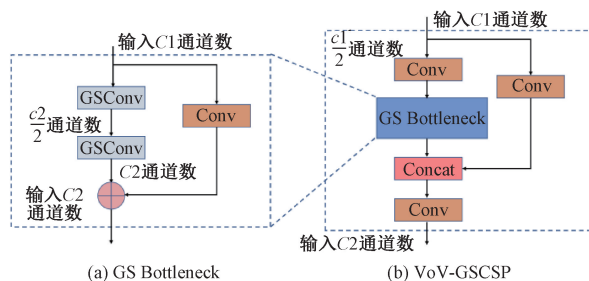


图 8 VoV-GSCSP 模块结构

Fig. 8 VoV-GSCSP module structure

3) WFU 网络结构

RT-DETR 通过简单的上采样和 Concat 级联操作实

现不同分辨率的特征融合。尽管这种方法简便,但会导致特征的高频细节与低频全局结构相互混叠,进而使细节信息被淹没,最终影响检测精度。针对这一问题,本文提出了 WFU 方法,该方法的核心在于利用小波变换的多尺度分析能力,对特征的低频全局信息与高频细节信息进行“解耦-针对性增强-重建融合”,从而更精确地整合不同尺度特征的有效信息^[19]。

WFU 利用小波变换(wavelet transform, WT)的多尺度分解特性,将特征拆分为低频子带和高频子带,随后针对这两类子带分别设计融合策略,最终通过小波逆变换(inverse wavelet transform, IWT)重建出高质量的融合特征。具体流程如图 9 所示。

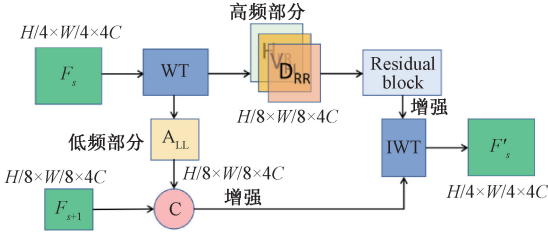


图 9 WFU 模块结构

Fig. 9 WFU module structure

首先对来自编码器的较大尺度特征 F_s 进行小波变换,得到与较小尺度特征 F_{s+1} 相同尺度的 4 个小波子带:

$$\{A_{LL}^s, H_{LR}^s, V_{RL}^s, D_{RR}^s\} = \text{WT}(F_s) \quad (8)$$

其中, A_{LL}^s 表示 F_s 的低频部分, $H_{LR}^s, V_{RL}^s, D_{RR}^s$ 表示 F_s 的 3 个高频部分。

小尺度特征 F_{s+1} 主要包含低频信息,将 A_{LL}^s 与其合并为增强的低频子带。同时,利用残差块对图像的高频成分进行增强,最终通过小波反变换得到输出 F'_s , 如式(9)所示。

$$F'_s = \text{IWT}(C(A_{LL}^s, F_{s+1}), R(H_{LR}^s, V_{RL}^s, D_{RR}^s)) \quad (9)$$

其中, IWT 为小波逆变换, C 为拼接, R 为标准残差块。这种融合方式有效的解决了 CCFM 中存在的高低频混叠问题。

2.4 损失函数优化

在目标检测任务中,边界框损失函数的性能显著影响模型训练的收敛速度与目标检测精度。原始 RTDETR 采用 GIoU 计算真实框与预测框的回归损失。尽管 GIoU 解决了两框不重叠时 IoU 为 0 的问题,但其对边界框形状变化过度敏感,易导致计算过程中产生较大损失偏差。在无人机航拍任务中,存在大量小目标,目标尺度变化大、背景复杂,因此本文提出 Focaler-ElIoU 与 NWD 加权融合损失函数改进方案。

Focaler IoU^[20] 聚焦于图像中小目标检测场景,通过融合 ElIoU 对边界框间距离、比例及纵横比差异的全面评估机制,有效缓解不规则目标形状对预测框的影响,提升目标定位精度。在 Focaler IoU 的边界框回归机制中,针对

VisDrone 数据集中目标的空间分布特性,将视野内占比大、特征明显的中大型目标(如近景车辆)归类为易样本,而将远距小目标(如远处行人、微小障碍物)定义为难样本。通过动态调整尺度因子 d 和 u (取值范围 $[0, 1]$),平衡训练过程中不同样本的信息挖掘难度,显著加速边界框回归的收敛进程。Focaler IoU 的具体计算过程如式(10)所示。

$$\text{IoU}^{\text{Focaler}} = \begin{cases} 0, & \text{IoU} < d \\ \frac{\text{IoU} - d}{u - d}, & d \leq \text{IoU} \leq u \\ 1, & \text{IoU} > u \end{cases} \quad (10)$$

ElIoU 损失包括 IoU 损失、中心距离损失、尺度损失和纵横比损失。损失计算过程如式(11)所示。

$$\begin{cases} L_{\text{ElIoU}} = 1 - \text{IoU} + D_{\text{loss}} + S_{\text{loss}} + R_{\text{loss}} \\ D_{\text{loss}} = \frac{d^2(c', c)}{d_{\text{max}}^2} \\ S_{\text{loss}} = \frac{|\omega' - \omega|}{\omega_{\text{max}}} + \frac{|h' - h|}{h_{\text{max}}} \\ R_{\text{loss}} = \left(1 - \frac{\omega'/h'}{\omega/h}\right)^2 \end{cases} \quad (11)$$

其中, c' 和 c 为预测框与真实框的中心坐标, $d(\cdot)$ 为欧氏距离, d_{max}^2 是覆盖两框的最小外接矩形对角线平方。 h', h 和 ω', ω 分别为两框的高和宽, ω_{max} 和 h_{max} 对应最大宽度和最大高度。ElIoU 通过同时考虑距离、尺度和形状差异,实现预测框与真实框的精准对齐,避免中心偏移。Focaler-ElIoU 边界框损失函数如式(12)^[21]所示。

$$L_{\text{Focaler-ElIoU}} = L_{\text{ElIoU}} + \text{IoU} - \text{IoU}^{\text{Focaler}} \quad (12)$$

Focaler-ElIoU 通过融合中心距离、尺度和形状约束,对无人机视角下的小目标检测具有良好的定位优化效果。但在处理极小规模目标(如像素面积 < 200)及密集分布场景时,仍存在尺度敏感导致的梯度失效和定位偏差问题,因此引入 NWD 以增强小目标的像素级定位鲁棒性。

Wang 等^[22]提出的 NWD 损失对目标尺度变化具有强鲁棒性,无需边界框重叠即可通过分布相似性度量实现有效回归,在小目标检测场景下效果显著。该方法将边界框建模为二维高斯分布,通过归一化 Wasserstein 距离刻画分布间差异。首先,对于水平边界框 $R = (C_x, C_y, h, \omega)$, 其中 (C_x, C_y) 为中心坐标, ω 和 h 分别表示宽高,将其建模为高斯分布 $N(\mu, \sigma)$, 具体建模形式为:

$$\begin{cases} \mu = \begin{bmatrix} \frac{C_x}{C_y} \end{bmatrix} \\ \sigma = \begin{bmatrix} -\frac{\omega^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \end{cases} \quad (13)$$

对于边界框 A 和 B,将其相似度转化为二维高斯分布的距离度量:设两分布分别为 $\mu_1 = N(m_1, \sigma_1), \mu_2 = N(m_2, \sigma_2)$, 则 μ_1 与 μ_2 之间的二阶 Wasserstein 距离可表

示为:

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \text{tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2^{1/2})^{1/2}) \quad (14)$$

对于由边界框 $A = (c_{x_a}, c_{y_a}, \frac{w_a}{2}, \frac{h_a}{2})$ 和 $B = (c_{x_b}, c_{y_b}, \frac{w_b}{2}, \frac{h_b}{2})$ 分别建模的高斯分布 N_a 和 N_b , 其二阶 Wasserstein 距离可简化为:

$$W_2^2(N_a, N_b) = \left\| \left(\begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\| \quad (15)$$

由于 $W_2^2(N_a, N_b)$ 属于距离度量, 无法直接作为相似性指标, 需通过归一化处理构建损失函数。最终的 NWD 损失表达式定义为:

$$\text{NWD}(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (16)$$

式中: C 为常量^[23]。

在 Focaler-EIoU 中引入 NWD 有效的形成优势互补, 增加了对小目标的鲁棒性, Focaler-EIoU 通过中心距离、宽高比、重叠度三钟损失项, 优化目标的整体形状匹配。NWD 补充微观精修从概率分布层面细化像素级定位, 尤其对极小规模目标的中心坐标偏移进行精准校正。因此最终得到的损失函数表达式为:

$$\text{loss} = \lambda[1 - \text{NWD}(N_a, N_b)] + (1 - \lambda)(1 - L_{\text{Focaler-EIoU}}) \quad (17)$$

式中: λ 为常数。

通过控制变量法测试权重系数 $\lambda = 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8$ 时的性能, 结果表明, 当 $\lambda = 0.7$ 时, mAP50 达到最大值 52.1%; 当 $\lambda \geq 0.75$ 时, mAP50 降至 51.9%, 且模型 FLOPs 增加 18%、推理速度下降 15%, 出现性能瓶颈与计算冗余; 而 $\lambda \leq 0.65$ 时, mAP50 均低于 52.0%。因此, 最终选择权重系数 $\lambda = 0.7$ 。

3 实验分析与结果

3.1 实验环境

实验基于 Windows 10 操作系统, 硬件配置包括 Intel(R) Xeon(R) Gold 5118 处理器(12 核 24 线程, 主频 2.3 GHz, 最大睿频 3.2 GHz), 内存为 64 GB, GPU 采用 NVIDIA Tesla P100(显存 16 GB)。软件环境使用 CUDA 11.6 和 PyTorch 1.13.1。实验未加载预训练模型权重, 训练轮次设定为 180, 批次大小为 8, 初始学习率设为 0.000 2, 所有输入图像统一调整为 640×640 pixels。

3.2 数据集

实验采用 VisDrone2019 数据集, 由天津大学机器学习与数据挖掘实验室 AISKEYEYE 提供, 总计包含 8 629 张图片, 分为训练集 6 471 张、验证集 548 张和测试集 1 610 张。

数据集涵盖 10 类目标, 包括行人、人员、自行车、汽车、面包车、卡车、三轮车、遮阳棚三轮车、公交车和摩托车。

3.3 评估指标

选择精确度(precision, P), 召回率(recall, R), 平均精度(arerage precision, AP), 平均精度均值(mAP), 模型参数量(Param)、千兆浮点计算量(GFLOPs)作为评价指标, 部分指标计算公式为:

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (18)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (19)$$

$$A_{\text{AP}} = \int_0^1 P(R) dR \quad (20)$$

$$A_{\text{mAP}} = \frac{1}{N} \sum_{i=1}^N A_{\text{AP}_i} \quad (21)$$

式中: N_{TP} 为模型正确检测到的正例样本数量; N_{FP} 为模型将负例预测为正例的样本数量; N_{FN} 为正例预测为负例的样本数量; N 为类别数量; A_{AP_i} 为第 i 个类别的 AP 值; mAP50 为 IoU 为 0.5 时的 mAP; mAP50:95 则是计算在 IoU 阈值从 0.5~0.95 内模型的 mAP。

3.4 消融实验

为验证各改进模块对模型性能的影响, 以 RT-DETR-R18 作为基线模型, 在 VisDrone2019 数据集上开展消融实验, 对不同模型的评价指标变化进行定量分析。为客观评估每个改进点, 将基准模型(RT-DETR-R18)定义为 A, 主干网络改进(Dfaster_net)定义为 B, 颈部网络改进 SWFM 定义为 C, 损失函数改进(NWD+Focaler EIoU)定义为 D。消融实验结果如表 2 所示。

由表 2 可知, 与基准模型 A 相比, 改进 B 在 mAP50 上提升 1.5%, 参数量减少 55%, GFLOPs 增加 14.8%; 改进 C 在 mAP50 上提升 0.3%, GFLOPs 下降 10%, 参数量下降 18.6%。当同时进行改进 B 和 C 时, mAP50 的提升幅度明显高于 B 与 C 单独提升幅度之和, 这表明改进 B 后输出的小目标特征更为精准, 为跨尺度特征融合提供了更高质量的输入素材; 改进 D 在 mAP50 上提升了 2.8%, 验证了改进损失函数对小目标定位精度的直接优化作用。

SwiftHawk-DETR 模型整合各模块后, 其精确度 P 、召回率 R 、mAP50 和 mAP50:95 相较于基线模型分别提升了 2.0%、3.3%、3.8% 和 3.0%, 且参数量降低 61%。与基准算法相比, 本文为提高小目标检测精度所做的改进不可避免地增加了计算复杂度, 使推理速度略有降低, 但计算量的增加(5.9 GFLOPs)处于可接受范围, 其带来的收益远高于成本。

综上所述, SwiftHawk-DETR 模型不仅在检测精度方面实现了显著突破, 还在模型轻量化方面获得了有效改进, 这使得该模型能够在实际工业部署中发挥重要作用。

表 2 在 VisDrone2019 数据集上的消融实验结果

Table 2 Ablation experimental results on the VisDrone2019 dataset

模型	$P/\%$	$R/\%$	mAP50/ $\%$	mAP50:95/ $\%$	Prama/ $(\times 10^6)$	GFLOPs
A	62.6	46.7	48.3	29.7	19.9	57.0
A+B	62.5	48.2	49.8	30.6	8.9	71.8
A+C	62.3	47.0	48.6	28.6	16.2	51.2
A+D	62.9	48.5	49.7	30.8	20.0	57.0
A+B+C	63.7	49.3	51.0	31.4	7.8	63.0
A+B+D	64.0	49.8	51.6	32.0	8.9	72.0
A+C+D	63.1	48.7	50.0	30.2	16.2	51.2
A+B+C+D	64.6	50.0	52.1	32.7	7.8	62.9

3.5 损失函数对比实验

为了评估改进后的损失函数性能,对损失函数进行了定量对比实验,通过与 GIoU、DlIoU、SIoU、ShapeIoU、NWD、EIoU 和 Focaler EIoU 的比较,结果如表 3 所示。实验表明,本文提出的方法获得了最高的 mAP 值,证实了改进损失函数的有效性。

表 3 不同损失函数性能对比

Table 3 Performance comparison of different loss functions (%)

损失函数	P	R	mAP50	mAP50:95
GIoU	62.6	46.7	48.3	29.7
DlIoU	62.5	46.8	48.6	29.7
SIoU	61.9	47.5	48.8	29.9
ShapeIoU	61.4	47.3	48.3	29.3
NWD	61.8	48.0	49.1	30.0
Focaler EIoU	62.0	48.2	49.2	30.4
本文	62.9	48.5	49.7	30.8

3.6 可视化分析

通过可视化分析可直观展现模型改进效果。本文将 SwiftHawk-DETR 模型与基线算法在 VisDrone2019 数据集上的检测结果进行了比较。

图 10 所示为改进前后的检测效果,其中上半部分为

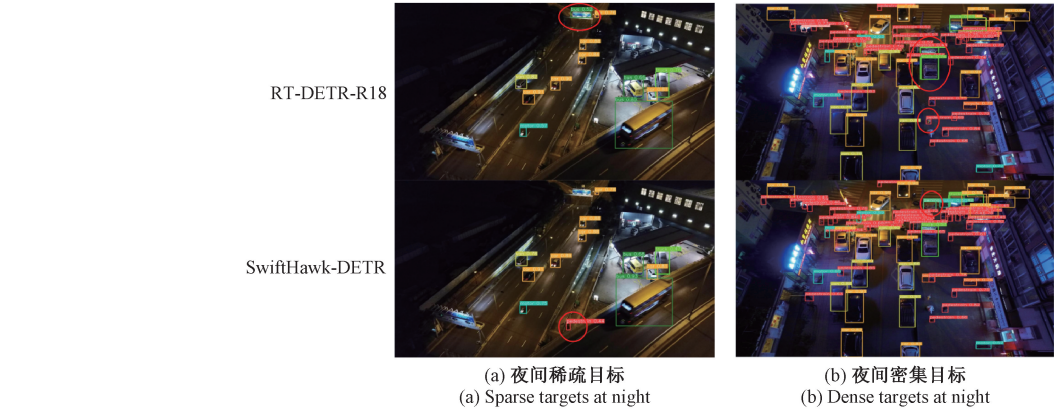
RT-DETR-R18 的检测结果,下半部分为 SwiftHawk-DETR 的检测结果。RT-DETR-R18 在图 10(a)中将广告牌错检成公交车且漏检一处行人;在图 10(b)中将一辆卡车重复检测,将路边的收起的遮阳伞错检成行人,漏检一辆三轮车;在图 10(c)中漏检遮挡下的汽车、三轮车上的的人;在图 10(d)中漏检摩托车和行人。

由于无人机视角下存在大量遮挡、模糊以及小目标,且检测类别存在外观相似的自行车、摩托车和三轮车,导致基线算法难以进行准确识别,进而出现误检、漏检和错检等问题。通过对比分析可知,SwiftHawk-DETR 模型在面对复杂场景、小目标占比多、遮挡严重的情况下表现更佳,体现出良好的稳定性,能更好地完成无人机目标检测任务。

此外,通过比较模型改进前后的混淆矩阵,可分析性能差异。对角线数值表示模型对样本的正确分类比例,而非对角线数值反映误检或漏检的比例。如图 11 所示,SwiftHawk-DETR 模型在图 11(b)中的对角线值显著高于图 11(a)中的 RT-DETR-R18 模型,表明其性能更优。

3.7 主流算法对比实验

为进一步评估 SwiftHawk-DETR 模型的性能优势,将其与主流算法进行对比实验,包括改进模型(Drone-YOLO-L^[24]、Deformable DETR^[6]、Sparse DETR^[25]、ESO-DETR^[26]、MGEF-DETR^[27]),同样选择 VisDrone 数据集,结果如表 4 所示。SwiftHawk-DETR 的 mAP50、



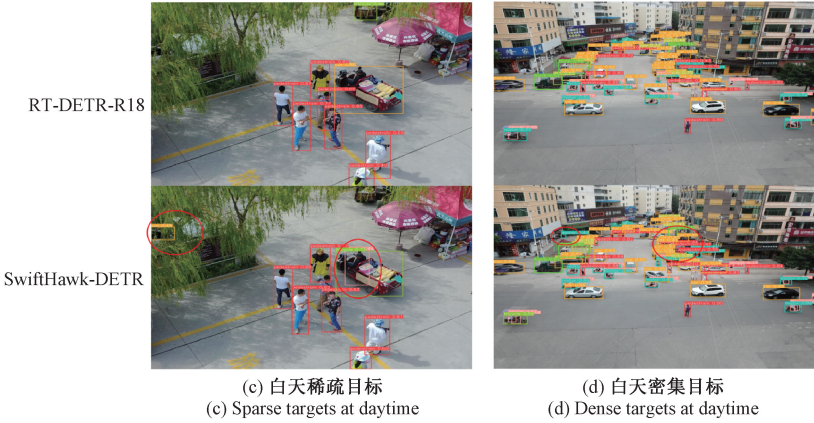


图 10 模型改进前后在 VisDrone2019 数据集上的检测结果对比

Fig. 10 Comparison of detection results on the VisDrone2019 dataset before and after model improvement

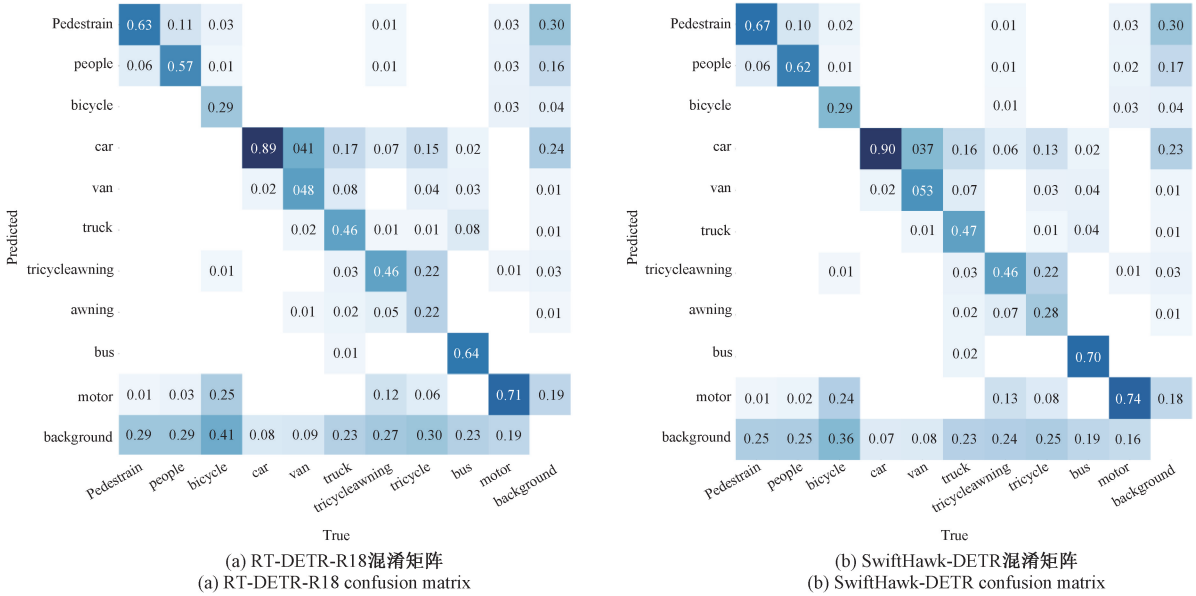


图 11 模型改进前后的混淆矩阵

Fig. 11 The confusion matrix before and after model improvement

表 4 不同模型在 VisDrone2019 数据集上的检测结果比较

Table 4 Comparison of various models' detection results on the VisDrone2019 dataset

算法	Pramas/($\times 10^6$)	mAP50:95/%	mAP50/%	GFLOPs	Inputsiz
Drone-YOLO-L	76.2	31.9	51.3	—	640×640
Deformable DETR	40.0	27.1	42.2	173.0	1 333×800
Sparse DETR	40.9	27.3	43.5	121.0	1 333×800
RT-DETR-R50	42.0	30.3	49.2	136.0	640×640
RT-DETR-R18	19.9	29.7	48.3	57.0	640×640
ESO-DETR	14.9	32.0	51.8	66.0	640×640
MGEF-DETR	17.2	31.9	51.3	68.2	640×640
SwiftHawk-DETR	7.8	32.7	52.1	62.9	640×640

mAP50:95 分别达到了 52.1%和 32.4%,在所有对比模型中达到最高,参数量为 7.8×10^6 ,在所有对比模型中达到最小。并且 SwiftHawk-DETR 与目前较先进的算法 ESO-DETR 和 MGEF-DETR 相比,mAP50 分别提高了 0.3%和 0.8%,mAP50:95 分别提高了 0.7%和 0.8%,参数量分别下降了 47.6%和 54.65%,计算量分别下降了 3.1 和 5.3 GFLOPs。显然,相较于 ESO-DETR 和 MGEF-DETR,SwiftHawk-DETR 的模型更为轻量,占用的存储空间和内存更少,推理速度更快,能耗更低,部署在边缘设备上的门槛也更低。此外,在更小、更快的基础上,SwiftHawk-DETR 实现了更高的 mAP 指标。这充分证明了 SwiftHawk-DETR 在模型性能上的优越性,使其成为面向实际工业部署的更优秀选择。

3.8 泛化实验

为了进一步验证所提算法的有效性和泛化性,本研究在 HIT-UAV 数据集上将 SwiftHawk-DETR 与当前多种主流目标检测算法进行了对比实验,结果如表 5 所示。显然,在 HIT-UAV 数据集上,SwiftHawk-DETR 的表现优于其他模型。与 YOLOv8m 相比,其 mAP50 和 mAP50:95 分别提升了 6.2%和 4.4%。相较于基线算法 RT-DETR-R18,其 mAP50 和 mAP50:95 分别提升了 4.2%和 3.8%,且其参数量仅为这两者的 30%和 41%。这表明 SwiftHawk-DETR 在保持较小模型体积和较低计算资源消耗的前提下,实现了更高的检测精度,在常见的视觉场景和复杂的红外设置中均表现出色,充分证明了其卓越的推广潜力,能够在挑战性条件下有效检测目标,进一步验

表 5 不同模型在 HIT-UAV 数据集上的检测结果比较

算法	<i>P</i> / %	<i>R</i> / %	mAP50 / %	mAP50:95 / %	Params / ($\times 10^6$)
YOLOv8m	82.0	76.5	79.8	51.0	25.8
YOLOv10m	85.5	71.0	78.2	47.6	16.4
RT-DETR-R18	83.6	78.8	81.8	51.6	19.9
RT-DETR-R50	87.5	78.6	82.5	52.0	42.0
SwiftHawk-DETR	85.2	80.7	86.0	55.4	7.8

证了其在多场景小目标检测任务中的适用性和鲁棒性。
图 12 直观展示了在不同场景下,SwiftHawk-DETR 的检测性能明显优于基准模型 RT-DETR-R18。其中,红

色椭圆框代表错检区域,绿色矩形框代表漏检区域,蓝色圆形框代表误检区域。在场景 1 中,SwiftHawk-DETR 正确识别出 RT-DETR 错检的其他车辆。在场景 2 和场景 3

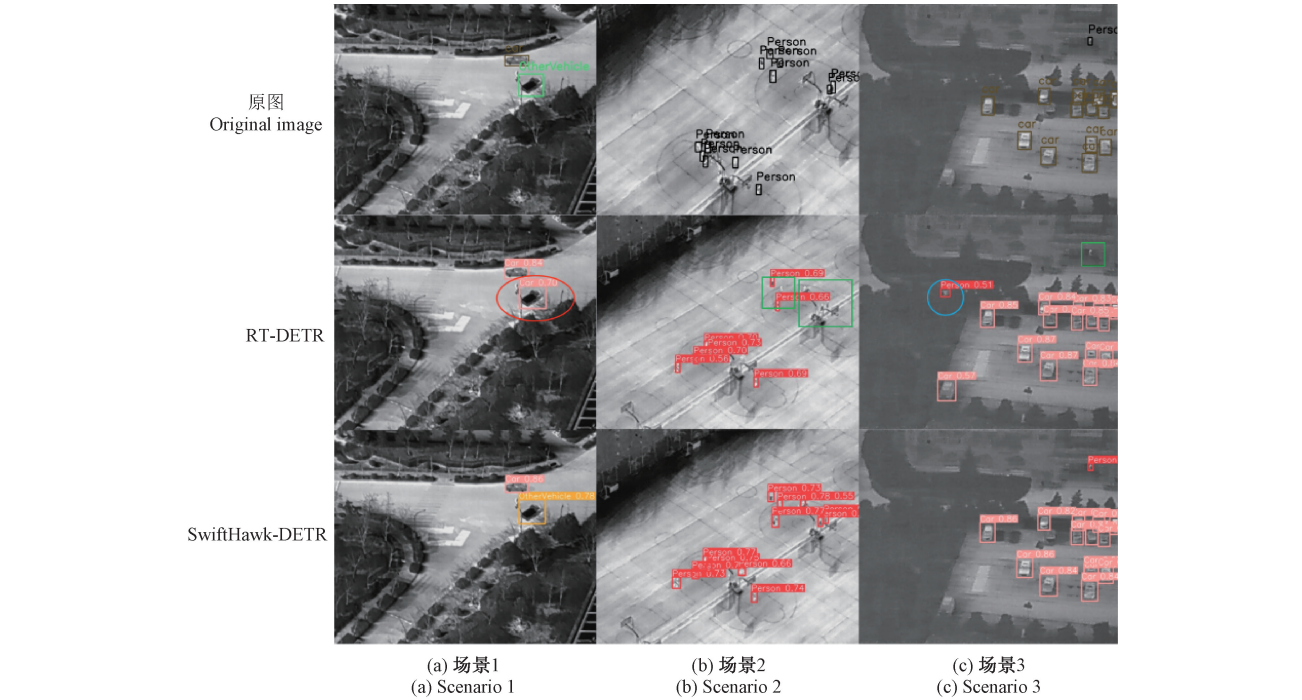


图 12 HIT-UAV 数据集上的检测结果对比

Fig. 12 Comparison of detection results on HIT-UAV datasets

中,SwiftHawk-DETR 成功识别出 RT-DETR 漏检的人。在场景 3 中,RT-DETR 出现将反光误检成人的情况,而 SwiftHawk-DETR 则有效避免了这一问题。

因此可以得出结论:即便在红外无人机视角下,SwiftHawk-DETR 依旧展现出优秀的检测性能、运行稳定性以及跨场景适应能力。在数据分布差异显著、环境条件多变的情形下,它仍能稳定维持高性能,这表明其设计思路适用于处理复杂场景中的无人机小目标检测。

4 结 论

本研究提出一种改进 RT-DETR 的高精度、低参数无人机小目标检测算法 SwiftHawk-DETR,该算法在无人机航拍检测任务中展现出优异性能。

首先,通过对特征金字塔的特征融合路径进行重构,新增微小目标检测层 P2 并移除大目标检测层 P5,在有效提升微小目标检测能力的同时,大幅减少网络参数。其次,设计 Dfaster_net 主干网络,削减计算冗余,并在深层引入 DCNv3 可变形卷积,增强上下文感知能力与几何特征建模能力,扩大感受野。然后,设计 SWFM 主干网络,通过引入 WFU,有效缓解上采样和 Concat 操作导致的边缘高频特征失真问题。引入 slimneck 架构,提高特征融合效率。最后,在检测头设计 Focaler EIoU+NWD 加权融合的损失函数,提高小目标检测精度与鲁棒性。

实验结果显示,本文算法在 VisDrone2019 和 HIT-UAV 数据集上取得了出色效果,多个指标表现良好,达成了算法轻量化、高精度的目标,领先于目前大多数主流目标检测模型,在无人机目标检测任务中具备更大优势。

此外,在后续研究中,将进一步改进模型轻量化,以适应计算资源有限的应用场景中的目标检测任务,顺应国产化发展趋势。

参考文献

- [1] 赵爽,黄怀玉,胡一鸣,等. 基于深度学习的无人机航拍车辆检测[J]. 计算机应用, 2019, 39(S2): 91-96.
Zhao Shuang, Huang Huaiyu, Hu Yiming, et al. Deep learning-based vehicle detection in UAV aerial photography[J]. Journal of Computer Applications, 2019, 39(S2): 91-96.
- [2] 张伟,庄幸涛,王雪力,等. DS-YOLO: 一种部署在无人机终端上的小目标实时检测算法[J]. 南京邮电大学学报(自然科学版), 2021, 41(1): 86-98.
Zhang Wei, Zhuang Xingtao, Wang Xueli, et al. DS-YOLO: A real-time small target detection algorithm deployed on UAV terminals[J]. Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2021, 41(1): 86-98.
- [3] Zhang J, Xie C M, Xu X, et al. A contextual

bidirectional enhancement method for remote sensing image object detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 4518-4531.

- [4] 张瑞芳,杜伊婷,程小辉. 无人机视角下小目标检测算法[J]. 激光与光电子学进展, 2025, 62(4): 0437002.
Zhang Ruifang, Du Yiting, Cheng Xiaohui. Small target detection algorithm from the perspective of UAVs[J]. Laser & Optoelectronics Progress, 2025, 62(4): 0437002.
- [5] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision-ECCV 2020. Lecture Notes in Computer Science. Cham: Springer, 2020: 213-229.
- [6] Zhu X Z, Su W J, Lu L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection[PP/OL]. V4. (2021-03-18). <https://arxiv.org/abs/2010.04159>.
- [7] Zhao Y A, Lyu W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2024: 16965-16974.
- [8] Lyu W Y, Zhao Y A, Chang Q Y, et al. RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer[PP/OL]. V1. (2024-07-24) [2025-10-25]. <https://arxiv.org/abs/2407.17140>.
- [9] 张晞,赖惠成,姜迪,等. MBFE-DETR: 多尺度边界特征增强下的无人机目标检测算法[J]. 计算机工程与应用, 2025, 61(17): 89-101.
Zhang Xi, Lai Huicheng, Jiang Di, et al. MBFE-DETR: A UAV target detection algorithm with multi-scale boundary feature enhancement[J]. Computer Engineering and Applications, 2025, 61(17): 89-101.
- [10] 向毅伟,蒋瑜,王琪凯,等. 多尺度特征优化的实时 Transformer 在无人机航拍中的研究[J]. 计算机工程与应用, 2025, 61(9): 221-229.
Xiang Yiwei, Jiang Yu, Wang Qikai, et al. Research on real-time Transformer with multi-scale feature optimization for UAV aerial photography [J]. Computer Engineering and Applications, 2025, 61(9): 221-229.
- [11] 程鑫森,张雪松,曹冰洁,等. 改进 RT-DETR 的小目标检测方法研究[J]. 计算机工程与应用, 2025, 61(15): 144-155.
Cheng Xinmiao, Zhang Xuesong, Cao Bingjie, et al. Research on improved small object detection methods

- for RT-DETR [J]. Computer Engineering and Applications, 2025, 61(15): 144-155.
- [12] Miao Q N, Wang T H, Wang R. Intelligent factory vehicle detection algorithm based on improved YOLOv8[J]. Instrumentation, 2025, 12(2), DOI: 10.15878/j. instr. 202500274.
- [13] 赵子琪, 李卫东, 李晓娟. 基于改进实时 Transformer 的航拍图像小目标检测算法[J]. 科学技术与工程, 2025, 25(13): 5527-5534.
- Zhao Ziqi, Li Weidong, Li Xiaojuan. Small target detection algorithm for aerial images based on improved real-time Transformer [J]. Science Technology and Engineering, 2025, 25(13): 5527-5534.
- [14] 苏佳, 杨梦凡, 张柏杨, 等. 改进 RT-DETR 的无人机小目标检测算法[J/OL]. 微电子学与计算机, 2024-1-13. <http://kns.cnki.net/kcms/detail/61.1123.tn.20241104.1032.004.html>.
- Su Jia, Yang Mengfan, Zhang Baiyang, et al. Improved RT-DETR-based small target detection algorithm for UAVs [J/OL]. Microelectronics & Computer, 2024-1-13. <http://kns.cnki.net/kcms/detail/61.1123.tn.20241104.1032.004.html>.
- [15] Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning [C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island: IEEE, 2023: 1-5.
- [16] 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 Yameng, Zhao Youquan, Sun Zhentao, et al. Construction of improved RT-DETR algorithm for detecting annular ripple defects of contact lenses[J]. Journal of Electronic Measurement and Instrument, 2024, 38(5): 1-9.
- [17] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks [C]//IEEE International Conference on Computer Vision. IEEE, 2017: 764-773.
- [18] Li H L, Li J, Wei H B, et al. Slim-neck by GSConv: A lightweight design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024, 21: 62.
- [19] Li W J, Guo H, Liu X N, et al. Efficient face super-resolution via wavelet-based feature enhancement network[C]//32nd ACM International Conference on Multimedia. ACM, 2024: 4515-4523.
- [20] Zhang H, Zhang S J. Focaler-IoU: More focused intersection over union loss [PP/OL]. V4. (2024-01-19)[2025-10-25]. <https://arxiv.org/abs/2401.10525>.
- [21] Zhang Y F, Ren W Q, Zhang Z, et al. Focal and efficient IoU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [22] Wang J W, Xu C, Yang W, et al. A normalized Gaussian Wasserstein distance for tiny object detection [PP/OL]. V2. (2022-06-14) [2025-10-25]. <https://arxiv.org/abs/2110.13389>.
- [23] 王迎龙, 孙备, 丁冰, 等. BG-YOLO: 复杂大视场下低慢小无人机目标检测方法[J]. 仪器仪表学报, 2025, 46(2): 255-266.
- Wang Yinglong, Sun Bei, Ding Bing, et al. BG-YOLO: Low-slow small UAV target detection method under complex large field of view[J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 255-266.
- [24] Zhang Z X. Drone-YOLO: An efficient neural network method for target detection in drone images[J]. Drones, 2023, 7(8): 526.
- [25] Roh B, Shin J W, Shin W, et al. Sparse DETR: Efficient end-to-end object detection with learnable sparsity [PP/OL]. V2. (2022-03-04) [2025-10-25]. <https://arxiv.org/abs/2111.14330>.
- [26] 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.
- [27] 侯林杰, 卢承方, 崔艳荣. MGEF-DETR: 多尺度门控增强融合的无人机目标检测算法[J/OL]. 电子测量技术, 2025-1-17. <https://link.cnki.net/urlid/11.2175.tn.20251117.1700.002>.
- Hou Linjie, Lu Chengfang, Cui Yanrong. MGEF-DETR: A UAV target detection algorithm with multi-scale gated enhancement fusion [J/OL]. Electronic Measurement Technology, 2025-1-17. <https://link.cnki.net/urlid/11.2175.tn.20251117.1700.002>.

作者简介

符强, 硕士, 正高级实验师, 主要研究方向为卫星导航、图像处理、SLAM。

E-mail: 1285389284@qq.com

彭章伟, 硕士研究生, 主要研究方向为目标检测。

E-mail: 1285389284@qq.com

纪元法(通信作者), 博士生导师, 主要研究方向为卫星导航、通信及数字信号处理。

E-mail: 2442319124@qq.com

任风华, 硕士, 主要研究方向为图像处理、信息处理。

E-mail: 1285389284@qq.com