

跨尺度特征融合的无人机航拍小目标检测算法^{*}

李孟欣 余长伟

(沈阳建筑大学电气与控制工程学院 沈阳 110000)

摘要: 针对无人机航拍影像中目标尺度多变、背景复杂所导致的小目标检测精度偏低和误检、漏检等问题,提出了一种基于 RT-DETR 改进的算法 CSFF-DETR。首先,在主干网络中设计上下文感知增强模块来提升主干网络对小目标特征的提取能力并抑制复杂背景干扰;其次,在颈部网络中通过引入高分辨率 P_2 特征层和设计跨尺度特征融合模块来实现不同尺度特征间的高效交互进而提高模型对小目标的检测能力;最后,设计多路径融合下采样模块来强化模型对小目标细节特征的保留与提取。在 VisDrone2019、Tinyperson 以及 UAVVaste 数据集上的实验表明,CSFF-DETR 相较于基准模型 RT-DETR, mAP_{50} 分别提升 3.6%、1.0%、2.8%; mAP_{50-95} 分别提升 3.0%、1.1%、1.6%,同时参数量降低 40.7%,满足了无人机平台对检测精度和轻量化的需求。

关键词: 无人机航拍影像;小目标检测;上下文感知;跨尺度融合;RT-DETR

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:** 520.6040

Cross-scale feature fusion algorithm for small object detection in UAV images

Li Mengxin She Changwei

(College of Electrical and Control Engineering, Shenyang Jianzhu University, Shenyang 110000, China)

Abstract: To address the issues of low precision, false detections and missed detections for small objects in UAV aerial images caused by large scale variations and complex backgrounds, an improved algorithm named CSFF-DETR based on RT-DETR is proposed. First, a context-aware enhancement module is designed in the backbone to improve small object feature extraction and suppress complex background interference. Second, a high-resolution P_2 feature layer is introduced into the neck network, along with a novel cross-scale feature fusion module, to enhance interaction between features at different scales and boost small object detection capability. Finally, a multi-path fusion downsampling module is designed to better preserve and extract fine-grained details of small objects. Experiments on the VisDrone2019, Tinyperson and UAVVaste datasets show that compared to the baseline RT-DETR model, CSFF-DETR achieves mAP_{50} improvements of 3.6%, 1.0% and 2.8% and mAP_{50-95} improvements of 3.0%, 1.1% and 1.6%, respectively, while reducing parameters by 40.7%. This meets the requirements for both detection accuracy and lightweight deployment on UAV platforms.

Keywords: UAV aerial images; small object detection; context awareness; cross-scale fusion; RT-DETR

0 引言

随着无人机在军事与民用领域的应用日益广泛,基于计算机视觉的目标检测技术成为实现无人机影像自动识别与定位的关键手段,对后续决策分析具有重要意义。然而,当前无人机航拍影像中存在目标尺度差异大、小目标居多、背景复杂等挑战^[1],且无人机的计算与存储资源受限。因此,如何在有限资源的约束下实现检测精度的最大化提升,

仍是当前研究的热点问题。

早期传统目标检测方法依赖手工设计特征和滑动窗口搜索,存在特征表达不足、计算效率低的问题。近年来,随着深度学习的发展,基于卷积神经网络(convolutional neural network, CNN)的目标检测方法逐渐成为主流,并广泛应用于无人机航拍影像目标检测中^[2],代表性算法包括两阶段的 R-CNN^[3]、Faster R-CNN^[4]和单阶段的 SSD^[5]、YOLO 系列等。这些方法虽简化了流程并提升了精度,但

仍存在明显局限。其中,两阶段算法虽检测精度高但计算开销大,难以部署于资源受限的无人机平台。单阶段算法虽推理速度快但受正负样本失衡的影响,在复杂场景下精度偏低,对小目标密集的航拍影像适应性较差。

为克服 CNN 方法的局限性,学界开始探索新型检测范式。其中,基于 Transformer 的端到端检测方法凭借其全局特征建模能力与序列到序列的学习机制,为无人机航拍影像小目标检测提供了新途径。代表性算法包括 DETR^[6]、Deformable DETR^[7] 及 RT-DETR (real-time detection transformer)^[8] 等。

为更好地适配无人机航拍影像小目标检测,RT-DETR 模型得到了多项针对性改进。Zhang 等^[9]提出的 UAV-DETR 通过带频率增强的多尺度特征融合模块融合多尺度下的空间域与频率域信息,实现对小目标细节特征的保留;Liu 等^[10]提出的 VRF-DETR 通过基于自适应感受野选择与空间注意力机制的多尺度上下文融合模块对多尺度特征进行动态校准,提升了模型的多尺度表征能力;Wang 等^[11]提出的 SF-DETR 通过双边交互式特征增强模块,实现层级特征间的双向信息流动,提升了密集集群中小目标的定位精度;Dong 等^[12]提出的 LKR-DETR 设计出了一种基于小波变换卷积的新型多尺度特征融合结构,能够有效利用低层特征图中的低频信息并且融合更多的原始信息;姜贺翔等^[13]提出的 EMRT-DETR 通过级联分组注意力机制优化倒置残差模块和特征交互模块,提升特征选择能力,实现目标检测信息的精细化获取;谢国波等^[14]提出的 Drone-DETR 通过设计结合并行分块感知注意力的主干网络,解决了因多次下采样操作导致的特征信息丢失问题;侯林杰等^[15]提出的 MGEF-DETR 通过设计基于自适应门控机制的多阶跨阶段门控聚合模块实现对小目标纹理特征的选择性增强;钟帅等^[16]提出的 MCS-DETR 在主干网络中设计多尺度边缘增强模块,通过构建边缘信息增强机制,并结合深度卷积操作和特征融合策略,实现不同尺度上提取特征信息。

尽管上述改进取得了不错的进展,但仍然存在一些局限。首先,这些方法在特征提取阶段仅聚焦目标自身特征,未建立目标与上下文信息的动态关联,导致在目标遮挡或复杂背景干扰下易出现误检、漏检;其次,现有多尺度融合策略多局限于相邻层级特征交互,非相邻层级特征间的信息传递受阻,难以有效捕捉小目标的细粒度特征与深层语义特征之间的互补关系。因此,针对上述不足,本文以 RT-DETR 为基准模型,结合无人机航拍影像特点,提出了跨尺度特征融合的无人机航拍影像小目标检测算法 CSFF-DETR (cross-scale feature fusion detection transformer),其创新点如下。

1) 提出了上下文感知增强模块 (context-aware enhancement module, CAE)。通过借鉴跨阶段局部网络的双路径结构设计,同时引入多膨胀率空洞卷积并行分支与池化-转置注意力分支协同建模小目标特征与其局部到

全局上下文信息之间的语义关联,增强主干网络在复杂背景下对小目标特征的提取能力。

2) 优化了颈部网络的多尺度特征融合机制。通过引入高分辨率 P_2 特征层和设计跨尺度特征融合模块 (cross-scale feature fusion module, CSFF),促进了浅层空间细节与深层语义信息的高效交互,有效缓解了因目标尺度多变和不同尺度特征间融合不充分导致的小目标检测精度偏低的问题。

3) 设计了多路径融合下采样模块 (multi-path fusion downsampling module, MPFD)。通过融合最大池化、平均池化以及深度可分离卷积池化等多种下采样方式,在压缩特征图空间尺寸的同时,最大化地保留了小目标的细节特征,提升了密集场景下小目标的检测性能。

1 RT-DETR 概述

RT-DETR 是基于 Transformer 架构的端到端目标检测模型,其架构主要包含 3 部分:以 ResNet18 为基础的主干网络,通过卷积层、池化层等提取图像的多尺度特征;作为核心组件的高效混合编码器,含基于注意力的同尺度特征交互模块 (attention-based intra-scale feature interaction, AIFI) 和基于 CNN 的跨尺度特征融合模块 (CNN-based cross-scale feature fusion, CCFF),二者协同处理多尺度特征以强化关联与融合;解码器则先通过 IoU 感知查询选择模块筛选初始对象查询,再经不确定性最小化查询选择机制优化查询质量,最终通过迭代优化生成目标的边界框及类别置信度。

2 CSFF-DETR 算法

为解决无人机航拍影像中因目标尺度多变、背景复杂所导致的小目标检测精度偏低和误检、漏检等问题,本文从主干网络的特征提取、颈部网络的多尺度融合以及下采样方式 3 个层面对 RT-DETR 进行改进。首先,在 RT-DETR 原有的主干网络中集成 CAE 模块来构建新的特征提取网络;其次,在颈部网络中引入高分辨率的 P_2 特征层和设计 CSFF 模块来改进 RT-DETR 原有的多尺度特征融合机制;最后,设计 MPFD 模块来替代 RT-DETR 原有的单一下采样方式。改进后的模型结构如图 1 所示。

2.1 上下文感知增强模块

针对无人机航拍影像中的小目标像素占比低、特征微弱且极易受复杂背景干扰的问题。本文提出一种 CAE 模块,其结构如图 2 所示。该模块借鉴了跨阶段局部网络 (cross stage partial network, CSPNet)^[17] 的双路径结构设计用以提升主干网络对小目标特征的提取能力。在此基础上,构建了一种局部与全局协同的上下文感知机制用以抑制复杂背景干扰,提升小目标与背景间的区分能力。

输入特征 $\mathbf{X}_i \in \mathbf{R}^{C \times H \times W}$ 经过 1×1 卷积进行特征对齐后,沿通道维度划分为两个分支。一条分支保留原始特征

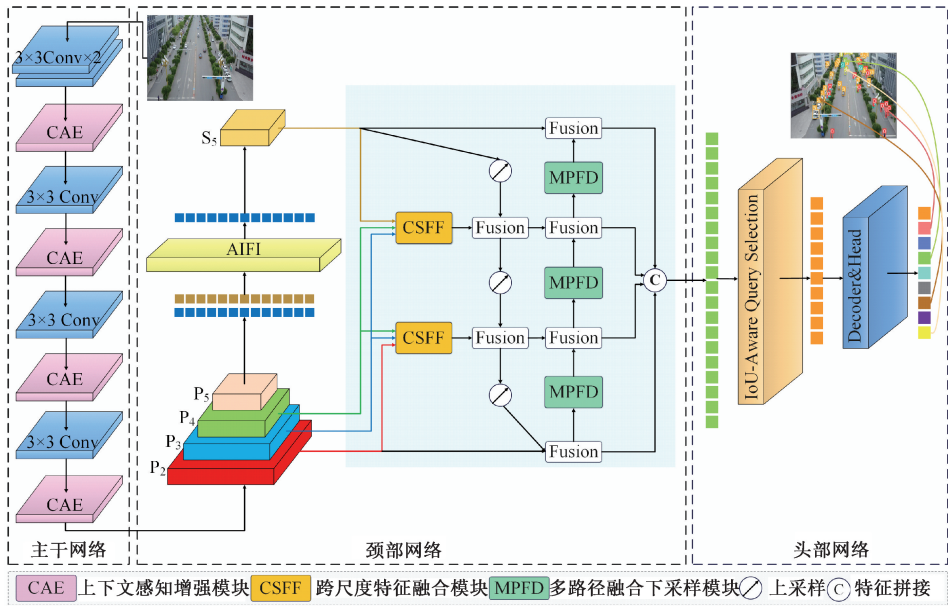


图 1 CSFF-DETR 模型架构
Fig. 1 CSFF-DETR model architecture

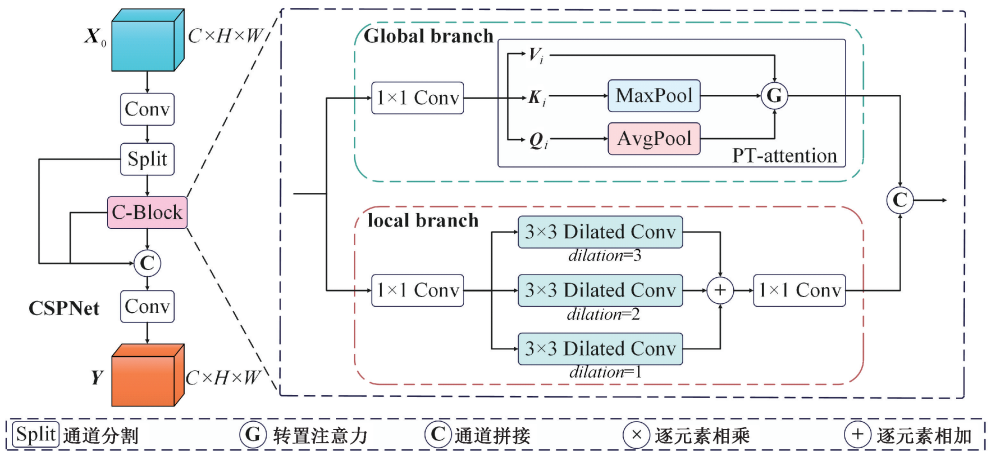


图 2 CAE 结构
Fig. 2 CAE architecture

并通过残差连接直接传递,另一分支则送入 C-Block 中进行深度特征变换。这种类似 CSPNet 的双路径设计能够在减少参数数量和计算量的同时显著缓解深层网络中的信息衰减问题,从而提升主干网络对于无人机航拍影像中小目标特征的提取能力。最后,两条路径的特征在通道维度上进行拼接与融合,得到 CAE 的输出 Y 。

C-Block 通过两个协同工作的分支构建了从局部到全局的上下文感知体系。其中,多膨胀率空洞卷积并行分支用以捕获小目标精细纹理特征并聚合其局部上下文信息;池化-转置注意力(PT-attention)分支用以建模小目标与全局上下文信息之间的长程语义依赖。首先,在局部上下文建模方面。空洞卷积^[18]是一种通过在卷积核元素间插入空洞(由膨胀率控制)来扩大感受野,且不增加参数数量的卷

积操作。基于此,本文设计的多膨胀率空洞卷积并行分支,部署了一组膨胀率(dilation)分别为 1、2、3 的 3×3 空洞卷积。根据感受野计算式(1)可知,这 3 个卷积在保持 3×3 参数数量的前提下,其有效感受野分别可近似于标准的 3×3 、 5×5 与 7×7 卷积。

$$R = d \times (k - 1) + 1 \quad (1)$$

其中, k 为卷积核大小, d 为膨胀率。感受野为 3×3 的卷积负责捕捉局部细粒度特征,而感受野更大的 5×5 和 7×7 卷积则负责提取中、远距离的语义信息。相较于大多数现有的改进方法直接并行使用多个大尺寸标准卷积核来扩大感受野,此设计能够在降低模型参数量与计算量实现模型轻量化的同时,系统性地捕获从细粒度到中、远距离的局部上下文信息。

然而,仅有局部上下文信息难以应对图像中全局性的复杂背景干扰。为此,本文引入池化-转置注意力(pooling ransposed attention, PT-attention)^[19]来捕获全局上下文信息。具体而言,PT-attention 首先分别对查询 Q_i 和键 K_i 施加平均池化与最大池化来进行全局范围内的特征聚合与筛选,这种并行双池化策略可有效地抑制背景噪声同时增强特征的稳定性。在此基础上,PT-attention 通过转置注意力 G 来高效建立全局范围内的长程语义依赖。如式(2)所示,G 首先计算池化后键的转置 $\bar{K}_i^T$ 与查询 $\bar{Q}_i$ 的矩阵乘积,经 Softmax 归一化后获得注意力权重,再通过矩阵转置操作调整注意力图的空间权重分布,最后与值 V_i 进行聚合。PT-attention 的计算公式为:

$$A_i^{\text{att}} = G(\bar{Q}_i, \bar{K}_i, V_i) = V_i \times [\sigma(\bar{K}_i^T \times \bar{Q}_i)]^T \quad (2)$$

式中: σ 表示 Softmax 操作, $\bar{K}_i$ 与 $\bar{Q}_i$ 分别为池化后的键和查询特征, V_i 为未经池化的原始值特征。通过上述

机制,PT-attention 能够高效整合全局范围内的语义信息,为区分小目标与复杂背景提供了关键的上下文依据。

2.2 改进的多尺度特征融合

1) 引入高分辨率 P_2 特征层

RT-DETR 及其大多数现有改进模型所采用的特征金字塔结构仅依赖 P_3 、 P_4 、 S_5 层级特征来应对多尺度目标检测,未能充分利用分辨率更高的浅层特征。这导致小目标的边缘与纹理等细节信息在特征提取过程中严重流失,从而制约了模型对于小目标的检测精度。

针对上述问题,本文通过引入 P_2 层级特征,扩充了特征金字塔的尺度范围。 P_2 层更高的空间分辨率(160×160 pixels)能够为小目标检测提供更丰富的细节特征,从源头上缓解因特征图分辨率不足导致的细节信息丢失问题。引入 P_2 特征层后的结构如图3所示。

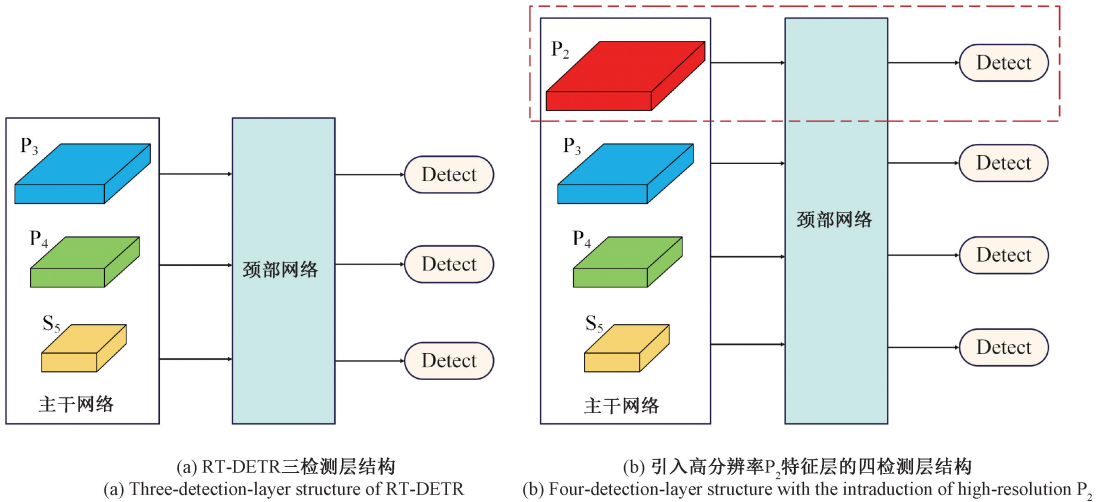


图3 引入高分辨率 P_2 特征层前后的结构对比

Fig. 3 Architecture comparison before and after integrating the high-resolution P_2 layer

2) 跨尺度特征融合模块

RT-DETR 的特征交互主要依赖于相邻层级,非相邻层级间的特征关联受限,阻碍了浅层空间细节与深层语义信息的充分交互。为此,本文设计了 CSFF 模块,其结构如图4所示。该模块包含多尺度特征对齐和通道与空间双重注意力加权融合两个阶段。

首先,通过 1×1 卷积将浅层特征 S 、中层特征 M 以及深层特征 D 统一至相同的通道数,并利用上采样与下采样操作将不同层级的特征对齐至相同空间尺寸。随后,将对齐后的特征沿通道维度拼接得到融合特征 $F_c \in \mathbf{R}^{3C \times H \times W}$ 。为突出特征融合后不同特征的重要性,对 F_c 施加全局平均池化压缩空间信息,再经 1×1 卷积与 Sigmoid 激活函数生成通道注意力权重 W_{ch} 。

$$W_{ch} = \text{Sigmoid}(\text{Conv}_{1 \times 1}(\text{GAP}(F_c))) \quad (3)$$

式中: $\text{Conv}_{1 \times 1}$ 为 1×1 卷积, GAP 为全局平均池化。

为进一步保留重要特征同时舍弃信息含量较低的特征,将 F_c 与 W_{ch} 逐元素相乘后再经 1×1 卷积调整通道数,得到通道加权后的特征 F_{ch} 。

$$F_{ch} = \text{Conv}_{1 \times 1}(W_{ch} \otimes F_c) \quad (4)$$

为聚焦于各层级特征共同关注的显著空间区域,分别将浅层特征 S_a 、中层特征 M_a 以及深层特征 D_a 通过 1×1 卷积压缩为单通道空间关注图,相加后经 Sigmoid 激活生成空间注意力权重 W_{sp} 。

$$W_{sp} = \text{Sigmoid}(\text{Conv}_{1 \times 1}(S_a) \oplus \text{Conv}_{1 \times 1}(M_a) \oplus \text{Conv}_{1 \times 1}(D_a)) \quad (5)$$

最终,利用空间注意力权重对通道加权特征进行聚焦,进一步优化特征表示,输出融合特征 F_f :

$$F_f = W_{sp} \otimes F_{ch} \quad (6)$$

通过上述基于注意力权重的加权融合机制,CSFF 模块能有效促进不同尺度特征间的高效交互,从而增强模型

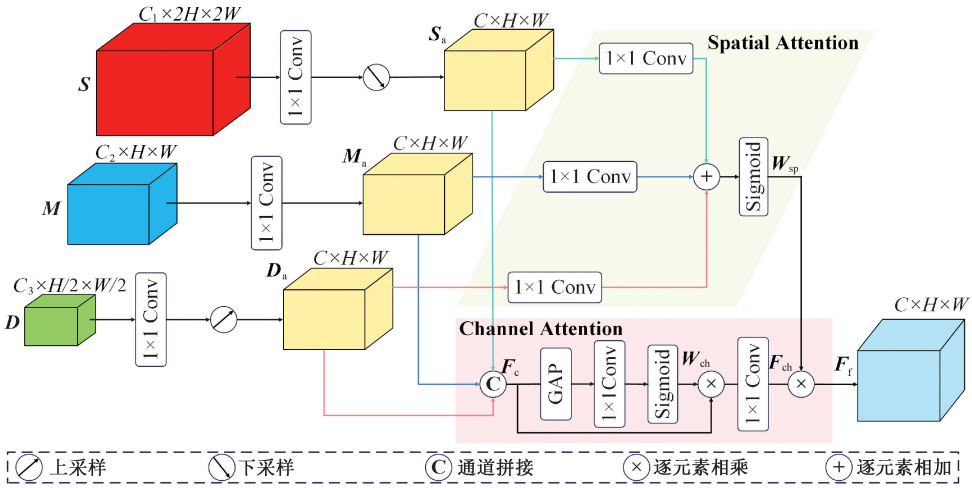


图 4 CSFF 结构

Fig. 4 CSFF architecture

在复杂背景下对小目标的检测能力。

2.3 多路径融合下采样模块

下采样作为特征提取的关键环节,直接影响特征图的空间分辨率与语义表征能力。RT-DETR 及其大多数现有改进模型均采用单一类型的卷积或池化进行下采样,虽计算高效,但却容易导致特征模糊甚至是关键目标信息的丢失。此问题在背景复杂且小目标居多的无人机航拍影像

中尤为突出。

针对上述问题,本文提出一种 MPFD 模块,其结构如图 5 所示。该模块创新性地将切片、卷积、最大池化以及平均池化 4 种各有不同侧重的下采样方式融合在了一起,使其在信息保留上形成互补,突破传统单一下采样方式的表征局限。在压缩特征图尺寸的同时,最大限度地保留信息的完整性与表达的多样性。

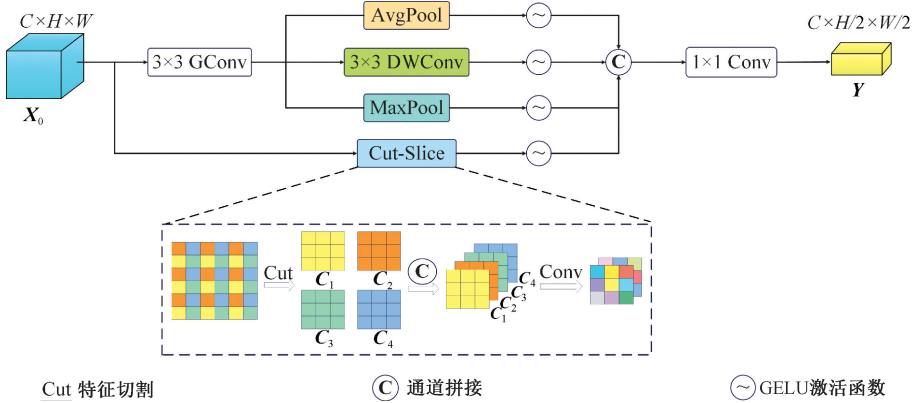


图 5 MPFD 结构

Fig. 5 MPFD architecture

输入特征 $X_0 \in \mathbb{R}^{C \times H \times W}$ 被划分为两个分支,一个分支经 3×3 分组卷积增强特征表达后被送入 3 条并行的下采样路径。路径 1 采用平均池化,通过局部均值压缩空间维度,能够有效抑制噪声并保留特征的全局分布信息;路径 2 采用深度可分离卷积,在降低计算量的同时避免了池化带来的纹理细节丢失,保留了更精确的空间结构;路径 3 采用最大池化,通过提取局部最大响应强化显著特征表达,增强边缘与角点等关键结构的表征。各分支计算流程为:

$$X_1 = \text{AvgPool}(\text{GConv}_{3 \times 3}(X_0)) \quad (7)$$

$$X_2 = \text{DWConv}_{3 \times 3}(\text{GConv}_{3 \times 3}(X_0)) \quad (8)$$

$$X_3 = \text{MaxPool}(\text{GConv}_{3 \times 3}(X_0)) \quad (9)$$

式中: $\text{GConv}_{3 \times 3}$ 为 3×3 分组卷积, $\text{DWConv}_{3 \times 3}$ 为 3×3 深度可分离卷积, AvgPool 为平均池化, MaxPool 为最大池化。

另一分支则引入切片下采样^[20]。该方法通过将特征图的相邻像素点分割并重组为 C_1 、 C_2 、 C_3 和 C_4 这 4 部分,再对不同像素点在通道维度上进行拼接。这种方法能建模不同子区域间的语义关联,弥补传统下采样操作中不同区域的特征在通道维度上缺乏交互的局限。

$$\mathbf{X}_4 = \text{Cut-Slice}(\text{GConv}_{3\times 3}(\mathbf{X}_0)) \tag{10}$$

式中:Cut-Slice 为切片下采样。

最后,将 4 条路径的输出特征在通道维度上进行拼接与融合,生成最终的下采样特征 $\mathbf{Y}$ 。

$$\mathbf{Y} = \text{Conv}_{1\times 1}(\text{Concat}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)) \tag{11}$$

式中:Concat 为通道拼接操作。

通过聚合上述不同路径的特征表示,MPFD 模块能够生成比任何单一路径都更具鲁棒性和丰富性的下采样结果,提升了对无人机影像中小目标细节特征的提取与保留能力。

3 实验结果与分析

3.1 实验环境

本文实验在 NVIDIA RTX4090(24 GB)GPU 上进行,采用 Python 3.8 编程语言及 PyTorch 1.13.1 深度学习框架。为保证对比公平性,所有对比模型均未使用预训练权重。训练配置如下:批量大小(batch size)设为 4,训练轮次(epoch)设为 300,输入图像尺寸为 640×640 pixels,初始学习率为 0.000 1,采用 AdamW 优化器进行模型优化,其余训练超参数均采用默认配置。

3.2 实验数据集介绍

VisDrone2019 数据集^[21]由天津大学机器学习与数据

挖掘实验室构建,是航拍影像目标检测领域主流基准数据集,含 8 629 张无人机航拍图像、10 个目标类别及约 260 万个标注样本,场景覆盖多种光照条件、天气状况等复杂环境场景。

TinyPerson 数据集^[22]专门为极小目标检测设计,聚焦于低分辨率航拍场景中的海上与陆地人类目标,含 1 610 张图像、72 651 个标注对象。

UAVVaste 数据集^[23]由波兹南理工大学机器人团队开发,是面向无人机环境垃圾检测的开源数据集,含 772 张校园航拍图像及 3 718 个标注框,其垃圾目标多为小尺寸。

3.3 评价指标

为全面评估本文所提目标检测模型的性能,实验采用 6 种评价指标:精确率(precision,P)、召回率(recall,R)、平均精确率(average precision,AP)、平均精确率均值(mean average precision,mAP)、模型参数量(Params)及计算量(FLOPs)。

3.4 消融实验

为验证本文提出的各部分改进的有效性,在 VisDrone2019 公开数据集上进行消融实验,实验以 RT-DETR-R18 为基准模型,逐一加入或替换本文所提出的改进方法,实验结果如表 1 所示。表 1 中 P_2 -CSFF 代表颈部网络多尺度特征融合部分的改进,其包含引入的高分辨率 P_2 特征层和跨尺度特征融合模块。

表 1 在 VisDrone2019 数据集上的消融实验结果

Table 1 Results of ablation experiments on the VisDrone2019 dataset

Baseline	CAE	P_2 -CSFF	MPFD	$P/\%$	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$	Params/ $(\times 10^6)$	GFLOPs
✓				62.1	46.8	48.1	29.5	19.9	57.0
✓	✓			62.6	48.6	50.2	31.1	13.8	44.8
✓		✓		62.5	48.8	50.4	31.4	18.0	72.6
✓			✓	62.6	47.3	48.7	30.2	19.8	63.2
✓	✓	✓		62.9	50.2	51.2	32.0	11.9	60.2
✓	✓		✓	63.0	49.5	50.9	31.3	13.7	48.9
✓		✓	✓	63.2	49.7	51.0	31.5	15.5	63.0
✓	✓	✓	✓	64.0	50.1	51.7	32.5	11.8	59.8

实验表明,单独引入 CAE 后,模型的 mAP_{50} 与 mAP_{50-95} 较基准模型分别提升 2.1%、1.6%,参数量与计算量分别降低 30.7%和 21.4%,表明其在增强对小目标特征提取能力的同时显著降低了模型复杂度,实现了模型的轻量化。单独引入 P_2 -CSFF 后,模型参数量小幅降低,虽然计算量小幅增加,但 mAP_{50} 与 mAP_{50-95} 较基准模型分别提升 2.3%和 1.9%,验证了其通过引入 P_2 特征层和促进不同尺度特征间的高效融合能够有效提升小目标检测精度。单独引入 MPFD 后,模型的各项指标较基准模型均有不同程度提升,表明 MPFD 通过融合多种下采样方式,可有效缓解传统单一下采样策略导致的特征丢失问题,为模型整体的检测性能做出了贡献。当同时引入 CAE 和 P_2 -

CSFF 后,模型的性能得到进一步提升,由于 CSFF-DETR 的主干网络中引入了轻量化的 CAE 模块,其在削减冗余参数的同时,可生成更具表征效率的多尺度特征。在此基础上, P_2 -CSFF 被设计用于融合 CAE 输出的、已优化的多尺度特征流,由于输入特征本身质量更高效, P_2 -CSFF 无需为冗余特征构建复杂的交互结构,自身结构得以简化,因此二者结合后进一步降低了模型参数量。当 3 个模块联合启用时,模型达到最优性能: P 、 R 、 mAP_{50} 和 mAP_{50-95} 较基准模型分别提升 1.9%、3.3%、3.6%、3.0%,参数量从基准模型的 19.9×10^6 降至 11.8×10^6 ,降幅达 40.7%,计算量仅小幅升高 4.9%。这充分体现了各模块设计的协同效率,实现了“目标检测精度有效提升”与“模型参数量

显著降低”的双重优化目标。

3.5 对比实验

为验证 CSFF-DETR 在无人机航拍影像小目标检测中的先进性,本文在 VisDrone2019 数据集上选取了覆盖不同检测架构、兼具领域代表性与技术时效性的算法展开对比:既包含经典两阶段检测算法(如 Faster-RCNN)、主流单阶段检测算法(包括最新发布的 YOLOv11m、YOLOv12m,以及针对小目标场景优化的 YOLO 变体 Mamba-YOLO、Hyper-YOLO),也纳入了

适配无人机航拍小目标检测场景的最新 RT-DETR 改进模型(如 SF-DETR、LKR-DETR 等)。通过覆盖两阶段、单阶段、Transformer(RT-DETR 改进模型)等多类检测架构的主流算法与最新方案,一方面可系统对比 CSFF-DETR 与不同类型检测方法的性能优劣;另一方面,通过与同场景适配的 RT-DETR 改进模型进行针对性对比,能够更充分地凸显本文改进模型在无人机航拍小目标检测任务中的技术优势与先进性。实验结果如表 2 所示。

表 2 在 VisDrone2019 数据集上的对比实验结果
Table 2 Results of comparative experiments on the VisDrone2019 dataset

模型	年份	$P/\%$	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$	$Params/(\times 10^6)$	GFLOPs
Faster-RCNN	2015	48.1	34.6	35.8	21.3	41.4	208.0
YOLOv8-m	2023	53.7	43.0	43.7	26.7	25.8	78.7
RT-DETR-R18	2023	62.1	46.8	48.1	29.5	19.9	57.0
YOLOv10-m ^[24]	2024	56.0	41.4	44.0	26.8	15.3	58.9
Mamba-YOLO ^[25]	2024	53.0	41.0	42.3	25.8	20.5	44.5
YOLOv11-m	2024	55.1	43.4	45.0	27.8	20.0	67.7
Hyper-YOLO ^[26]	2025	51.7	40.1	41.1	24.9	13.5	33.8
YOLOv12-m	2025	54.0	42.8	44.3	27.2	19.6	59.5
SF-DETR	2025	63.3	49.9	51.0	31.8	14.7	55.7
LKR-DETR	2025	62.9	48.5	49.8	30.6	17.7	61.7
VRF-DETR	2025	—	—	51.4	31.8	13.5	44.3
EMRT-DETR	2025	62.0	46.6	48.8	30.5	13.8	68.1
Drone-DETR	2025	62.6	48.7	50.4	31.1	19.1	68.8
MGEF-DETR	2025	63.6	49.8	51.3	31.9	17.2	68.2
CSFF-DETR(本文)	2025	64.0	50.1	51.7	32.5	11.8	59.8

在检测精度方面,CSFF-DETR 整体表现最优: P 、 R 、 mAP_{50} 、 mAP_{50-95} 分别为 64.0%、50.1%、51.7%、32.5%,均位列所有对比模型首位。相较于传统单阶段与两阶段算法,CSFF-DETR 的 P 、 R 、 mAP_{50} 及 mAP_{50-95} 均有显著提升,体现出了 Transformer 架构的端到端目标检测算法在无人机航拍影像目标检测上的优势。与基准算法 RT-DETR 和最新改进算法 SF-DETR、LKR-DETR、MGEF-DETR 等相比,CSFF-DETR 的 mAP_{50-95} 分别提升了 3.0%、0.7%、1.9%和 0.6%,验证了本文所提出的改进方法对提升小目标检测精度的有效性和先进性。

在模型效率方面,CSFF-DETR 的参数量仅为 11.8×10^6 ,所有对比模型中最低,较 Hyper-YOLO 降低 12.6%;计算量为 59.8 GFLOPs 与单阶段算法 YOLOv12m 相当,且显著低于传统两阶段算法。实验结果表明,CSFF-DETR 在检测精度与模型轻量化之间实现了良好平衡。

3.6 可视化分析

为直观验证 CSFF-DETR 的检测性能,本文以 RT-DETR 为基准,从 VisDrone2019 数据集的密集遮挡、高空

小目标、夜间低光照 3 类典型场景中选取具有代表性的图像进行可视化对比。图 6(a)~(c)中从左至右依次为 RT-DETR 检测图、局部对比放大图和 CSFF-DETR 检测图。其中,RT-DETR 检测图中以圆形框标注漏检、三角形框标注误检,CSFF-DETR 检测图中以矩形框标注正确检测区域。

图 6(a)所示密集遮挡场景中,RT-DETR 将斑马线安全岛误检为车辆,并漏检被遮挡的边缘车辆;图 6(b)所示高空小目标场景中,RT-DETR 误将通风装置识别为货车,并漏检停车位上的车辆;图 6(c)所示夜间低光照场景中,RT-DETR 对角落人群漏检较多,并将暗区树叶误检为车辆。相比之下,CSFF-DETR 在上述 3 类典型复杂场景中均实现了更准确的检测。

3.7 泛化实验

为进一步验证 CSFF-DETR 在不同数据集上的泛化能力,本文选取公开数据集 TinyPerson 与 UAVVaste 开展泛化实验,为确保对比的全面性与公平性,选取了代表不同技术路线的先进模型作为基准,包括最新通用检测器(YOLOv11m、YOLOv12m)、面向小目标的专用轻量模

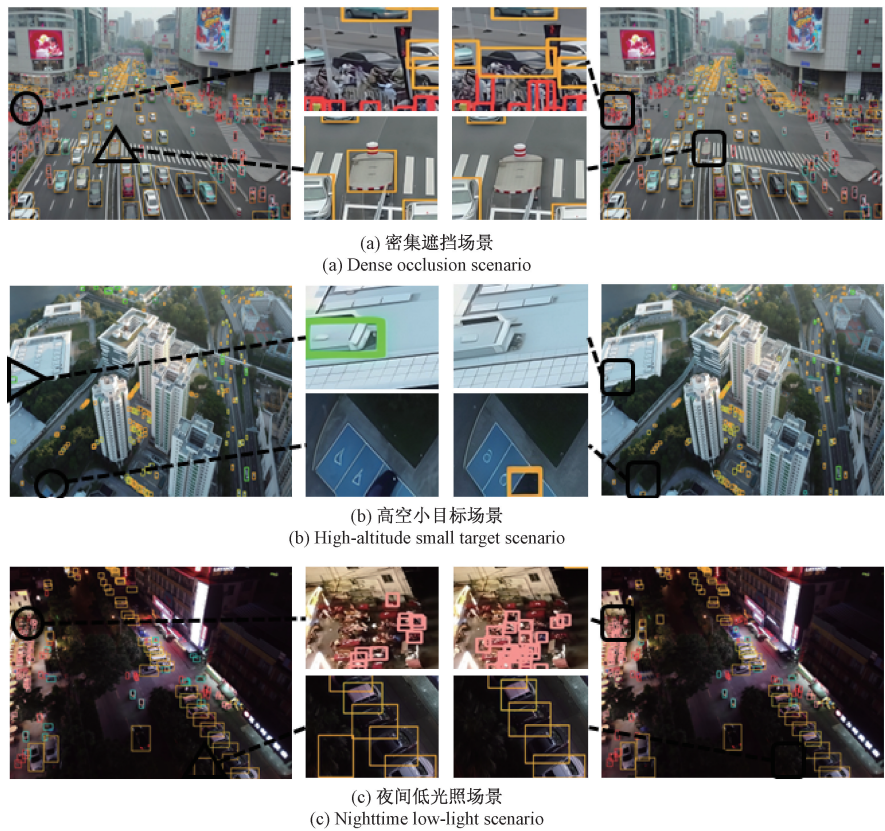


图 6 典型场景下的可视化对比

Fig. 6 Visual comparison diagrams under typical scenarios

型(Hyper-YOLO)、基于新兴架构的模型(Mamba-YOLO)以及 Transformer 范式下的高效实时检测器(RT-DETR-R18)。这一对比设置覆盖了当前目标检测领域的主流研究方向,能够客观评估 CSFF-DETR 在无人机航拍小目标检测场景下的泛化性能与技术优势。

实验结果如表 3 所示。在检测精度方面,CSFF-

DETR 在 TinyPerson 数据集上的 mAP_{50} 、 mAP_{50-95} 分别为 28.3%、10.8%,在 UAVVaste 数据集上的 mAP_{50} 、 mAP_{50-95} 分别为 82.7%、50.5%,均优于所有对比模型;在模型效率方面,CSFF-DETR 的参数数量同样低于其他对比模型,计算量与单阶段算法 YOLOv12m 相当,实现了检测精度提升与模型效率优化的良好平衡。

表 3 在 Tinyperson 和 UAVVaste 数据集上的对比实验结果

Table 3 Results of comparative experiments on the Tinyperson and UAVVaste datasets

数据集	模型	$P/\%$	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$	Params/ $(\times 10^6)$	GFLOPs
Tinyperson	YOLOv11m	42.0	20.9	21.3	7.9	20.0	67.7
	YOLOv12m	42.8	20.3	20.6	7.5	19.6	59.5
	Hyper-YOLO	41.6	19.5	20.1	7.3	13.5	33.8
	Mamba-YOLO	40.8	20.2	20.2	7.3	20.5	44.5
	RT-DETR-R18	48.4	28.8	27.3	9.7	19.9	57.0
	CSFF-DETR(本文)	46.8	30.8	28.3	10.8	11.8	59.8
UAVVaste	YOLOv11m	88.7	69.1	76.2	44.2	20.0	67.7
	YOLOv12m	86.7	68.4	76.3	43.8	19.6	59.5
	Hyper-YOLO	83.0	70.4	75.8	43.6	13.5	33.8
	Mamba-YOLO	85.6	67.8	75.1	42.8	20.5	44.5
	RT-DETR-R18	91.3	75.1	79.9	48.9	19.9	57.0
	CSFF-DETR(本文)	91.6	75.3	82.7	50.5	11.8	59.8

此外,与 VisDrone2019 和 Tinyperson 数据集相比,尽管 UAVVaste 数据集的数据量更少,但改进模型依然取得了较高的检测精度,表明 CSFF-DETR 的优异性能并不依赖大规模的标注数据,其架构设计本身具有较强的特征学习与泛化能力。

4 结 论

针对无人机航拍影像中目标尺度多变、复杂背景干扰导致的小目标检测精度偏低和误检、漏检等问题,本研究提出一种跨尺度特征融合的无人机航拍影像小目标检测算法 CSFF-DETR。通过在主干网络中集成上下文感知增强模块、在颈部网络中引入高分辨率 P_2 特征层并设计跨尺度特征融合模块以及采用多路径融合下采样模块代替原模型单一的下采样方式等一系列针对 RT-DETR 的改进,显著提升了小目标检测性能并优化了模型效率。经在 VisDrone2019、Tinyperson 以及 UAVVaste 数据集上的实验验证,证明了改进模型在无人机航拍影像小目标检测中的有效性和先进性。

未来工作将聚焦网络剪枝与知识蒸馏技术,进一步压缩模型体量与计算开销,以适配资源受限的无人机平台,推动低空遥感检测技术的实际应用。

参考文献

- [1] 苏雨蕾, 黄丹丹, 刘智, 等. 无人机对地小目标检测方法研究[J]. 电子测量与仪器学报, 2024, 38(9): 144-154.
- [2] 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.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [4] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]//14th European Conference on Computer Vision, 2016: 21-37.
- [6] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]//16th European Conference on Computer Vision, 2020: 213-229.
- [7] 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.
- [8] ZHAO Y, 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.
- [9] ZHANG H X, LIU K, GAN ZH X, et al. UAV-DETR: Efficient end-to-end object detection for unmanned aerial vehicle imagery[C]//2025 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2025: 15143-15149.
- [10] LIU W B, SHI L R, AN G CH. An efficient aerial image detection with variable receptive fields [J]. Remote Sensing, 2025, 17(15): 2672.
- [11] WANG H T, GAO J W. SF-DETR: A scale-frequency detection transformer for drone-view object detection[J]. Sensors, 2025, 25(7): 2190.
- [12] DONG Y, XU F CH, GUO J H. LKR-DETR: Small object detection in remote sensing images based on multi-large kernel convolution[J]. Journal of Real-Time Image Processing, 2025, 22(1): 1-14.
- [13] 姜贺翔, 司占军, 王晓喆. 改进 RT-DETR 的无人机图像目标检测算法[J]. 计算机工程与应用, 2025, 61(1): 98-108.
- [14] 谢国波, 黎逍, 林志毅. 面向无人机影像目标检测的 Drone-DETR 算法[J]. 电光与控制, 2025, 32(10): 70-76.
- [15] XIE G B, LI X, LIN ZH Y. Drone-DETR algorithm for drone image object detection [J]. Electronics Optics & Control, 2025, 32(10): 70-76.
- [16] 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]. 电子测量技术, 2026, 46(9): 177-191.
- [17] HOU L J, LU CH F, CUI Y R. MGEF-DETR: Multi-scale gated enhancement fusion for UAV object detection algorithm [J]. Electronic Measurement

- Technology, 2026, 46(9): 177-191.
- [16] 钟帅, 王丽萍. MCS-RETR: 改进 RT-DETR 的无人机航拍图像目标检测方法[J]. 航空学报, 2025, 46(22): 331987-331987.
- ZHONG SH, WANG L P. MCS-RETR: An improved RT-DETR-based target detection method for UAV aerial images [J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(22): 331987-331987.
- [17] 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: 1571-1580.
- [18] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. ArXiv preprint arXiv: 1511.07122, 2016.
- [19] NOMAN M, FIAZ M, CHOLAKKAL H, et al. ELGC-Net: Efficient local-global context aggregation for remote sensing change detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-11.
- [20] LU W, CHEN S B, TANG J, et al. A robust feature downsampling module for remote-sensing visual tasks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-12.
- [21] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop, 2019: 213-226.
- [22] YU X H, GONG Y Q, JIANG N, et al. Scale match for tiny person detection [C]//2020 IEEE Winter Conference on Applications of Computer Vision, 2020: 1246-1254.
- [23] KRAFT M, PIECHOCKI M, PTAK B, et al. Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle[J]. Remote Sensing, 2021, 13(5): 965.
- [24] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real time end-to-end object detection [J]. ArXiv preprint arXiv: 2405.14458, 2024.
- [25] WANG Z Y, LI CH, XU H Y, et al. Mamba YOLO: A simple baseline for object detection with state space model [J]. ArXiv preprint arXiv: 2406.05835, 2024.
- [26] FENG Y F, HUANG J G, DU SH Y, et al. Hyper-YOLO: When visual object detection meets hypergraph computation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 2388-2401.

作者简介

李孟歆, 博士, 教授, 硕士研究生导师, 主要研究方向为目标检测、图像处理、智能控制。

E-mail: limengxin@sjzu.edu.cn

余长伟(通信作者), 硕士研究生, 主要研究方向为计算机视觉、目标检测。

E-mail: shechangweihn@163.com