DOI:10.19651/j.cnki.emt.2519618

MGEF-DETR:多尺度门控增强融合的无人机目标检测算法*

侯林杰[1,2] 卢承方[1,2] 崔艳荣[1,2]
(1.长江大学计算机科学学院 荆州 434023;2.长江大学人工智能科研平台 荆州 434023)

摘 要:无人机航拍图像中的小目标检测面临目标尺寸微小、背景干扰复杂、特征表达不充分等关键技术挑战。针对现有 RT-DETR 模型在小目标特征提取和多尺度融合方面的局限性,提出一种自适应多尺度门控增强融合检测模型(MGEF-DETR)。该方法通过设计多阶跨阶段门控聚合模块(MCGA),通过自适应门控机制实现小目标纹理特征的选择性增强;构建 Micro-OmniPyramid 小目标特征金字塔,集成 SPD 卷积稀疏编码和跨阶段增强空间核模块(CESK),建立小目标特征的无损传递通路;引入增强特征关联模块 EFC,通过分组注意力和多级重建策略优化跨尺度特征融合;设计内部修正惩罚距离 IoU 损失函数(IMIoU),增强边界回归对小目标的敏感性。在 VisDrone2019 数据集上的实验结果表明,MGEF-DETR 相比基线模型 RT-DETR 在 mAP@0.5 和 mAP@0.5:0.95 指标上分别提升 3.9%和 3.1%,同时参数量减少 13.6%。在 TinyPerson 和 CODrone 数据集上的验证进一步证实了算法的泛化能力,表明该方法在保持轻量化的同时显著提升了航拍场景下小目标检测的精度和效率。

关键词:无人机目标检测;RT-DETR;小目标;多尺度特征融合;门控机制
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

MGEF-DETR: Multi-scale gated enhancement fusion for UAV object detection algorithm

Hou Linjie[1,2] Lu Chengfang[1,2] Cui Yanrong[1,2]
(1. School of Computer Science, Yangtze University, Jingzhou 434023, China;
2. Artificial Intelligence Research Platform, Yangtze University, Jingzhou 434023, China)

Abstract: Small object detection in UAV aerial imagery encounters critical challenges including extremely small target sizes, complex background interference, and insufficient feature representation. Addressing the limitations of existing RT-DETR models in small object feature extraction and multi-scale fusion, this paper proposes an adaptive multi-scale gated enhancement fusion DETR (MGEF-DETR). A multi-order cross-stage gated aggregation (MCGA) module is designed to achieve selective enhancement of small object texture features through adaptive gating mechanisms. A Micro-OmniPyramid feature pyramid is constructed by integrating space-to-depth (SPD) convolution sparse encoding and cross-stage enhanced spectral kernel (CESK) modules, establishing lossless transmission pathways for small object features. An enhanced feature correlation (EFC) module is introduced to optimize cross-scale feature fusion through grouped attention and multi-level reconstruction strategies. An inner-modified penalty distance IoU (IMIoU) loss function is designed to enhance boundary regression sensitivity for small objects. Experimental results on the VisDrone2019 dataset demonstrate that MGEF-DETR achieves improvements of 3.9% and 3.1% in mAP@0.5 and mAP@0.5:0.95 metrics respectively compared to the baseline RT-DETR, while reducing parameters by 13.6%. Validation on TinyPerson and CODrone datasets further confirms the generalization capability of the algorithm, indicating significant improvements in both accuracy and efficiency for small object detection in aerial scenarios while maintaining lightweight characteristics.

Keywords: UAV object detection; RT-DETR; small object detection; multi-scale feature fusion; gated mechanism

0 引 言

随着无人机技术的飞速发展,其在农业、航拍、快递运输以及监控等诸多领域的应用日益广泛。无人机搭载的摄像头能够收集大量的视觉数据,而对这些数据进行自动理解与分析,尤其是目标检测任务,成为计算机视觉领域的一

收稿日期:2025-08-17
*基金项目:国家自然科学基金面上项目(62077018)资助

个重要研究方向[1]。目标检测作为计算机视觉的基础任务之一，其目的是识别和定位图像中的目标对象，广泛应用于自动驾驶、安防监控、医疗影像分析等众多领域[2]。

目前，计算机视觉中的目标检测技术取得了长足发展，从早期的两阶段检测模型到后来的单阶段模型，出现了众多经典算法和架构[3]。两阶段模型以 R-CNN[4-5] 系列为代表，通过区域提议和分类两步实现高精度检测，但计算复杂度较高；单阶段模型则以 YOLO[6]、SSD[7] 等为代表，将检测过程整合为一步端到端预测，速度更快但最初精度相对偏低。随着 RetinaNet[8] 引入 Focal Loss 缓解了一阶段检测正负样本不平衡问题，单阶段检测器的精度劣势被显著缩小。此后，YOLO 系列算法不断演进在精度和速度上取得平衡，成为实际应用的主流选择。同时，Transformer[9] 架构开始应用于目标检测领域，Carion 等[10] 提出的 DETR (detection transformer)模型通过序列建模和一体化的端到端训练省去了 NMS 后处理，展示了新的检测范式。DETR 在小目标和训练收敛速度方面存在一定不足，Deformable DETR 等改进模型通过多尺度特征融合等技术提升了对小目标的检测性能。Zhao 等[11] 提出的 RT-DETR(real-time detection transformer)模型作为一种新兴的实时目标检测模型，融合了 Transformer 机制与高效编码解码结构，成为首个实时端到端目标检测器，由于省去非极大值抑制(non-maximum suppression，NMS)步骤从而进一步提升了推理效率。

然而，在实际应用中，小目标检测仍然是一个极具挑战性的问题。小目标通常具有尺寸小、特征不明显、易受背景干扰等特点，导致传统方法在准确识别和定位小目标方面存在局限性。为了提高小目标的检测精度，吴一全等[12] 系统综述了近年无人机小目标检测的进展，指出通过增强多层次特征的表达和融合，可以有效缓解小目标检测精度低的问题。邓天民等[13] 则提出了特征复用与重组策略，在减少模型参数的同时生成更多特征图以关注小目标语义信息。对于单阶段检测器，谢椿辉等[14] 针对无人机影像中小目标尺度多样且易受环境干扰的问题，提出了 Drone-YOLO 模型，通过增加小目标检测的多尺度检测分支，提升了模型对小目标的多尺度感知能力，但该方法在一定程度上增加了计算复杂性。刘洋等[15] 针对小目标检测中局部信息丢失和漏检率高的问题，提出了 LDF-YOLO 模型，通过引入 ConvFFN 局部特征增强模块以及 LCBHAM 特征转换模块，有效提升了小目标的检测精度和召回率，但该方法仍基于传统 YOLO 架构，在处理复杂航拍场景中的全局上下文建模能力有限。贺智轩等[16] 为改善无人机航拍图像中目标尺寸微小、多尺度变化显著以及复杂场景干扰导致的检测精度不足，提出了 DMF-YOLOv11 模型，设计了双重双向特征金字塔以及多分支混合卷积模块，有效提升模块针对小目标和密集场景的检测性能。以 Transformer 架构为基础的检测器方面，亦有研究者对小目标场景进行了有针对性的优化。胡佳乐等[17] 通过改进 RT-DETR 模型提出 DyASF 特征融合结构，设计动态尺度序列特征融合和三重特征编码模块，有效减少因上下采样导致的小目标特征信息丢失，但该方法在处理极端尺度变化场景时多尺度特征融合策略仍需优化。姜贸翔等[18] 提出基于 SimAM 注意力与倒置残差的 EMRT-DETR，通过增强浅层特征提取能力提升小目标检测精度，但该方法在引入多个改进模块后检测速度有所降低。Liu 等[19] 提出一种轻量化的 ESO-DETR 算法，通过设计 GSHA 注意力主干模块以及 EMASlideVariFocal 损失函数，降低了模型复杂度，但在处理长尾分布的困难样本时还有较大提升空间。

近年来，因 Transformer 架构基于全局信息交互进行特征学习，能有效地对图像的全局依赖关系进行建模，在目标检测领域展现出比传统 CNN 更优越的性能。受此启发，本文提出一种基于 RT-DETR-R18 改进的轻量化网络模型 MGEF-DETR。此方法将轻量级多阶跨阶段门控聚合模块(multi-order cross-stage gated aggregation，MCGA)结构作为主干网络，设计小目标特征金字塔 Micro-OmniPyramid，结合增强特征关联模块(enhanced feature fusion with channel-wise attention，EFC)构建高效的多尺度特征融合架构，并采用边界框损失函数(inner-modified penalty distance iou，IMIoU)以达到在保持轻量化的同时提升小目标检测精度的目的。最后，在 VisDrone2019[20]、TinyPerson[21] 及 CODrone[22] 无人机航拍数据集上实验验证了所提方法在航拍小目标检测中的有效性。

1　MGEF-DETR

RT-DETR 设计了一种高效混合编码器及带辅助预测头的解码器构成的实时端到端目标检测架构，通过解耦局部尺度内的特征交互与全局跨尺度的特征融合，以高效建模多尺度特征表示，引入 IoU 感知查询选择机制优化解码器初始化，经迭代优化后输出目标类别与边界框。本文提出一种基于 RT-DETR-R18 改进的轻量化网络模型 MGEF-DETR，其结构如图 1 所示。

MGEF-DETR 在 RT-DETR 架构基础上进行了系统性改进。输入一张原始分辨率的航拍图像，MGEF-DETR 采用轻量级的 MCGA 结构作为主干网络提取多尺度特征图。随后为特征图添加内容与位置查询编码并送入 Micro-OmniPyramid 特征金字塔，通过跨尺度特征增强通路设计实现小目标细粒度特征的有效保留。在特征融合阶段，EFC 模块替代传统拼接操作，通过自适应权重分配和语义关联建模增强多层次特征融合效果。利用 IoU 感知查询选择机制筛选有效特征，送入解码器进行特征解码，并通过 IMIoU 损失函数优化边界框回归，提升模型对小目标对象的定位准确度，最后通过辅助预测头输出检测框的位置与类别。

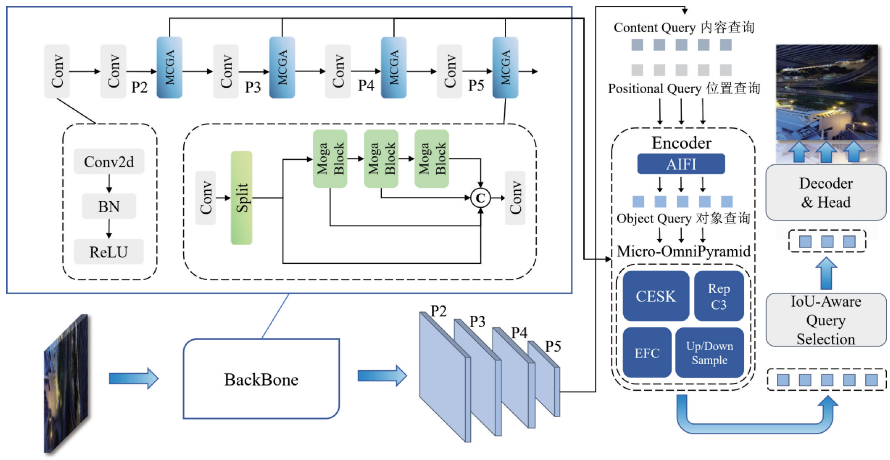
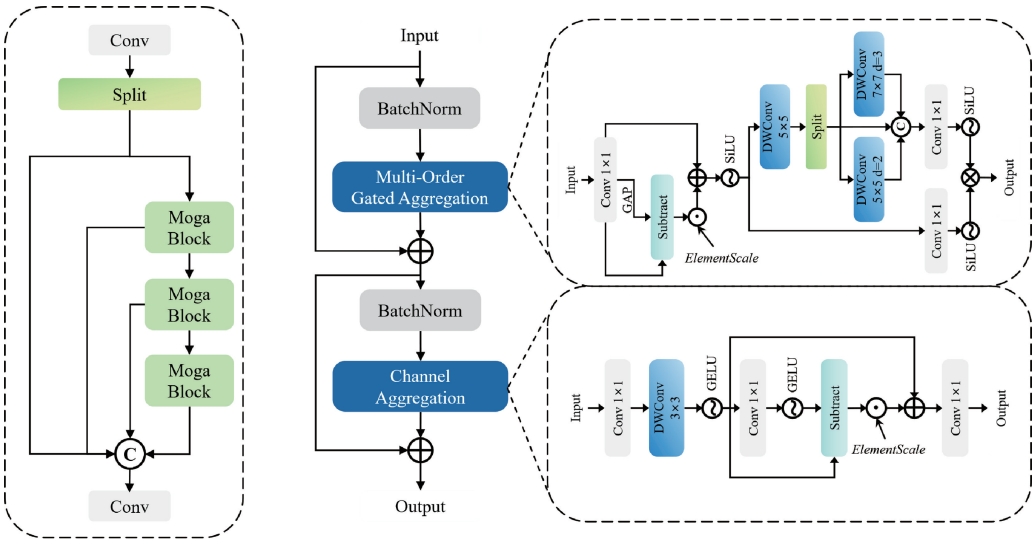


图 1 MGEF-DETR 模型结构

Fig. 1 The structure of the MGEF-DETR model

1.1 轻量级多阶跨阶段门控聚合模块 MCGA

针对 RT-DETR-R18 远程空间依赖捕获能力受限,在高分辨率视觉内容与复杂语义场景建模中存在固有的局部感知域约束等问题,从 MogaNet[23]中的多阶门控聚合机制(multi-order gated aggregation, MGA)得到启发,提出轻量级跨阶段聚合增强模块 MCGA,以此为基础采用分层特征传递机制重构主干网络架构。MCGA 通过多阶深度卷积处理不同尺度的语义上下文,耦合跨阶段局部网络(cross stage partial network, CSPNet)的梯度流优化机制[24]与 MogaNet 的自适应门控策略,以促进跨尺度特征的有效融合,建立全局-局部协同感知机制,为模型的检测性能提供更具判别性的特征表示。如图 2 所示,MCGA 模块保留了 CSPNet 的并行结构设计思想,并用多阶门控网络(Multineck)作为其内部的瓶颈结构。对于输入特征 $\boldsymbol{X} \in \mathbf{R}^{B\times C\times H\times W}$,模块首先通过 Conv 层和 Split 操作将特征分为主分支和跳跃分支,其中主分支 $\boldsymbol{X}_1 \in \mathbf{R}^{B\times C/2\times H\times W}$ 经过 n 个串联的 MogaBlock 处理,跳跃分支 $\boldsymbol{X}_2 \in \mathbf{R}^{B\times C/2\times H\times W}$ 直接传递,最终通过 Concat 操作拼接后经过 Conv 层输出 $\mathbf{R}^{B\times C\times H\times W}$ 维度特征。MGA 作为核心组件,替代传统卷积块的单一特征提取操作,实现空间-通道双路径协同感知模式。



(a) MCGA模块结构
(a) The module structure of MCGA

(b) MogaBlock模块结构
(b) The module structure of MogaBlock

图 2 MCGA 网络结构

Fig. 2 The network structure of MCGA

多阶深度可分离卷积层通过并行的 5×5 和 7×7 深度卷积核构建低阶、中阶和高阶的层次化特征交互机制。具体地,输入特征按照[1:3:4]的比例分割为 3 个子特征,分别经过扩张率为[1,2,3]的深度卷积处理:

$$\boldsymbol{Y}_{\text{low}},\boldsymbol{Y}_{\text{mid}},\boldsymbol{Y}_{\text{high}} = \text{Split}(\boldsymbol{X},[C/8,3C/8,C/2]) \tag{1}$$

$$\boldsymbol{Y}'_{\text{low}} = \text{DWConv}^{d=1}_{5\times5}(\boldsymbol{Y}_{\text{low}}) \tag{2}$$

$$\boldsymbol{Y}'_{\text{mid}} = \text{DWConv}_{5\times5}^{d=2}(\boldsymbol{Y}_{\text{mid}}) \tag{3}$$

$$\boldsymbol{Y}'_{\text{high}} = \text{DWConv}_{7\times7}^{d=3}(\boldsymbol{Y}_{\text{high}}) \tag{4}$$

该递进式膨胀策略使模型能够在不同感受野范围内同时捕获多粒度的特征响应模式。处理后的多阶特征通过通道维度的拼接操作进行融合，最后通过逐点卷积实现通道间的信息交互和特征整合。该机制在空间维度上首先通过 GAP 操作实现全局信息聚合，然后通过 *ElementScale* 参数 σ 进行特征调节：

$$x = x + \sigma(x - \text{GAP}(x)) \tag{5}$$

其中，σ 为可学习的参数，初始值设定为 1×10^{-5}，通过残差连接强化局部-全局特征的互补性。随后，门控聚合策略通过双分支结构实现自适应特征选择：

$$\text{output} = \text{SiLU}(\text{Conv}_{1\times1}(\text{gate}(x)))\odot \text{SiLU}(\text{Conv}_{1\times1}(\text{value}(x))) \tag{6}$$

该门控机制通过逐元素乘法实现特征的自适应权重分配，有效增强了特征表达能力。

在通道维度上，ChannelAggregation 通过特征分解策略实现通道信息的有效重分配。该模块首先通过 Conv 1×1 进行通道扩展，随后利用 DWConv 3×3 提取局部特征，经过 GELU 激活函数处理后再通过 Conv 1×1 降维，最后通过 Subtract 操作和可学习的 *ElementScale* 缩放因子 σ 调控分解强度：

$$x = x + \sigma(x - \text{GELU}(\text{Conv}_{1\times1}(\text{DWConv}_{3\times3}(\text{Conv}_{1\times1}(x))))) \tag{7}$$

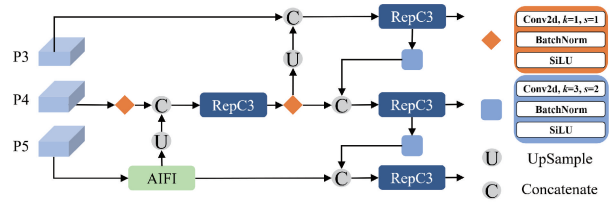
从计算复杂度角度分析，MCGA 模块的主要计算开销包括多阶深度卷积的浮点运算量 $O(C\cdot H\cdot W\cdot k^2)$ 和门控聚合机制的复杂度 $O(2\cdot C\cdot H\cdot W)$，其中 k 为卷积核大小。相比传统 ResNet-18 BasicBlock，MCGA 模块通过深度可分离卷积和跨阶段部分连接设计有效控制参数增长，保持了较低的计算复杂度。

改进主干网络通过多阶特征交互和自适应门控聚合的设计，解决了传统卷积网络在小目标特征提取方面的局限性。该网络架构在保持计算效率的前提下，有效提升了模型对航拍图像中小目标的特征感知能力，有效抑制了背景噪声的干扰，增强了目标与背景的特征区分度。

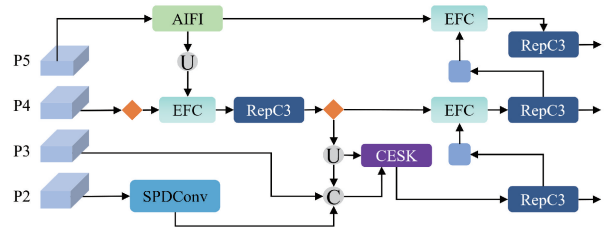
1.2　小目标特征金字塔 Micro-OmniPyramid

在航拍无人机小目标检测任务中，传统特征金字塔网络(feature pyramid network，FPN)[25]在上采样过程中容易导致小目标的边界、轮廓和纹理等关键细节丢失或模糊化。如图 3(a)所示，RT-DETR 模型采用跨阶段特征融合(cnn-based cross-scale feature fusion，CCFF)结构通过上采样和拼接操作构建多尺度特征，但仍存在浅层高分辨率信息利用不足的问题。为改善上述问题，提出一种名为 Micro-OmniPyramid 的特征融合架构。

不同于直接引入 P2 检测层[18]增加计算开销的方法，如图 3(b)所示，该模块通过并行处理 3 路不同尺度的特征并在通道维度进行融合：来自 P2 层的高分辨率特征经 SPD 卷积稀疏编码策略[26]进行空间到深度转换，该模块采用步长为 2 的下采样策略后接核大小为 3×3、步长为 1 的卷积层将通道数从 64 扩展至 128；P4 特征(256 通道)通过 1×1 卷积降维至 128 通道后经 2 倍上采样恢复至 P3 对应的空间尺度；P3 特征(128 通道)则通过跨层连接直接传递。3 路特征在通道维度拼接形成 384 通道的多尺度融合特征图，随后输入 CESK 进行跨阶段特征增强，最后通过 RepC3 模块完成特征提取与通道重构，输出增强特征。



(a) CCFF结构
(a) The structure of CCFF



(b) Micro-OmniPyramid结构
(b) The structure of Micro-OmniPyramid

图 3　特征融合网络结构

Fig. 3　The network structure of feature fusion

1.3　跨阶段增强空间核模块 CESK

CESK 构建局部-中尺度-全局三层级感受野体系，实现对不同尺度特征的自适应权重分配。在特征融合阶段，原始小目标特征图与经过多尺度处理的特征图通过残差连接方式进行加权融合，最后经过降维操作完成特征空间的优化重构。

如图 4(a)所示，CESK 设计思想来源于为 OmniKernel 特征提取模块[27]，采用跨阶段部分连接架构降低计算复杂度。在 CESK 结构中，输入特征首先经过 1×1 卷积进行通道变换并保持通道数不变，然后按照通道分割比例划分为两个分支：25%通道进入 OmniKernel 处理分支进行多尺度特征提取，75%通道作为恒等映射分支直接传递。两个分支在通道维度拼接后通过 1×1 卷积完成特征融合。

如图 4(b)所示，OmniKernel 核心模块采用分解重构策略将原始卷积核解耦为四种异构深度卷积核：点域感受野的 DConv 1×1、垂直感受野的 DConv 31×1、标准感受野的 DConv 31×31、水平感受野的 DConv 1×31。通过并行处理实现多方向感受野的全向覆盖。输入特征首先经过 1×1 卷积和 GELU 激活进行通道映射，4 个异构卷积分支和全局感知模块的特征通过加权求和机制进行融合，经 ReLU 非线性激活和逐点卷积降维处理后，最终通过残差

机制与输入特征进行自适应融合,与输入特征通过残差连接进行自适应融合。

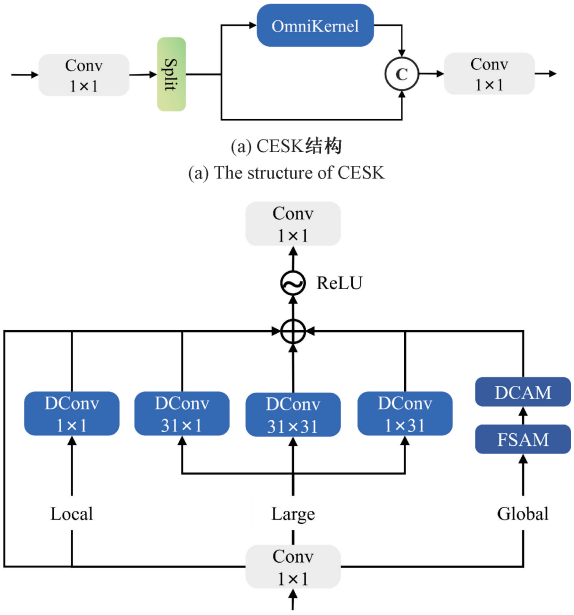


图 4 CESK 模块

Fig. 4 The module of CESK

如图 5 所示,全局感知模块包含双域通道注意模块(dual-domain channel attention module, DCAM)与频率空间注意模块(frequency-spatial attention module, FSAM)。DCAM 通过频域通道注意力(frequency channel attention, FCA)和空域通道注意力(spatial channel attention, SCA)的级联实现双域特征增强,其处理过程可表示为:

$$\boldsymbol{X}_{\mathrm{FCA}} = \left|\mathcal{F}^{-1}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{GAP}\left(\mathcal{F}\left(\boldsymbol{X}\right)\right)\right)\odot\mathcal{F}\left(\boldsymbol{X}\right)\right)\right| \tag{8}$$

$$\boldsymbol{X}_{\mathrm{SCA}} = \mathrm{Conv}_{1\times1}\left(\mathrm{GAP}\left(\boldsymbol{X}_{\mathrm{FCA}}\right)\right)\odot\boldsymbol{X}_{\mathrm{FCA}} \tag{9}$$

其中,$\mathcal{F}$ 和 $\mathcal{F}^{-1}$ 分别表示二维快速傅里叶变换(fast Fourier transform, FFT)和逆快速傅里叶变换(inverse fast Fourier transform, IFFT),GAP 表示全局平均池化,$|\cdot|$ 表示复数的模长运算。

FSAM 在 $\boldsymbol{X}_{\mathrm{SCA}}$ 基础上采用双路径结构:一路通过 1×1 卷积进行通道调制,另一路通过 1×1 卷积后进行 FFT 变换到频域,两路特征在频域进行复数域乘法运算,再通过 IFFT 变换回空域并取模长,最终通过可学习参数 α 和 β 完成残差加权融合。

引入 Micro-OmniPyramid 模块的优势在于其多层级并行处理结构,不仅提高了网络全局范围内对多尺度特征的理解和表达能力,而且通过局部模块的结构优化,确保模型在不过度增加计算开销的情况下,能够更高效地捕捉并学习跨尺度的特征信息,增强模型在小尺度场景下的表征能力。

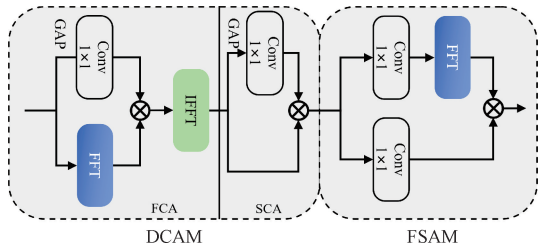


图 5 DCAM 与 FSAM 模块

Fig. 5 The modules of DCAM and FSAM

1.4 特征融合模块 EFC

RT-DETR 针对特征交互设计的 CCFF 结构主要基于特征拼接 Concat 操作实现不同尺度特征图的整合与交互,实现将主干网络输出的多层级特征空间维度统一,建立跨分辨率特征的直接连接通路,但该操作本质上仍属于静态融合范畴,未能充分建模特征通道间的语义依赖关系,在应对航拍场景中背景干扰强烈、小目标特征微弱且易被淹没的复杂情况时,其融合效果仍存在优化空间。

为改善上述问题,引入增强特征融合策略 EFC[28],该策略专门设计用于两个不同尺度特征图间的自适应融合,通过替换原有特征融合架构中部分特征拼接操作,实现多尺度特征的自适应融合和语义信息的有效保持。EFC 特征融合模块如图 6 所示。

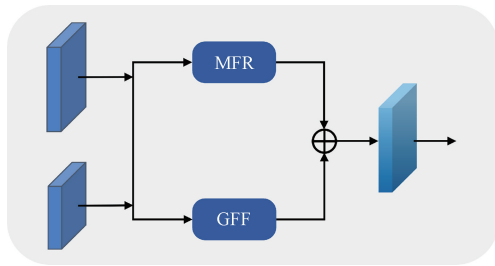


图 6 EFC 模块

Fig. 6 The module of EFC

EFC 策略的核心创新在于构建了一个双分支自适应特征融合框架,该框架集成了分组特征聚焦单元(group-wise feature focusing unit, GFF)和多级特征重建模块(multi-level feature reconstruction module, MFR)。GFF 单元通过对融合特征进行通道分组并应用多层次分组注意力机制思想,捕获异构语义层次的特征激活模式;MFR 模块则利用自适应权重分配动态调控多尺度特征的贡献,并通过深度可分离卷积与逐点卷积的组合实现特征的渐进式重构。该设计在保持原始特征语义一致性的同时,有效增强了跨尺度特征间的互补性,从而为微小目标物体提供了更丰富、判别性更强的特征表示。

如图 7 所示,GFF 单元通过将融合后的特征图按通道维度均匀划分为 G 个子组,每个子组独立进行特征增强处理。具体而言,对于输入特征 $\boldsymbol{X}_{\mathrm{global}}\in\mathbf{R}^{C\times H\times W}$,首先通过全局注意力门控机制生成空间注意力权重,然后将特征图分

解为 $\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_G\}$，其中 $G=4$。对于第 g 个特征子组 $\boldsymbol{X}_g\in\mathbf{R}^{C/G\times H\times W}$，GFF 单元基于自适应特征聚合的处理过程可以表示为：

$$\mathcal{A}_g^{(t)}=\mathrm{Softmax}\left(\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{X}_g^{(i,j)}\odot\Phi\left(\frac{1}{HW}\sum_{k=1}^{HW}\boldsymbol{X}_g^{(k)}\otimes\boldsymbol{W}_\alpha^{(g)}\right)}{\sqrt{d_k}+\epsilon}\right)\circ\mathcal{T}_\beta(\boldsymbol{X}_g)\tag{10}$$

其中，$\mathcal{A}_g^{(t)}$ 表示第 g 个子组在时间步 t 的动态注意力响应矩阵，$\Phi(\cdot)$ 为非线性激活函数，$\boldsymbol{W}_\alpha^{(g)}\in\mathbf{R}^{(C/G)\times(C/G)}$ 是第 g 组的可学习参数矩阵，$\otimes$表示 Kronecker 积运算，$\circ$表示 Hadamard 积积，$\mathcal{T}_\beta(\cdot)$ 为基于层归一化的变换函数，$d_k=C/G$ 为缩放因子，ϵ 为数值稳定性常数。

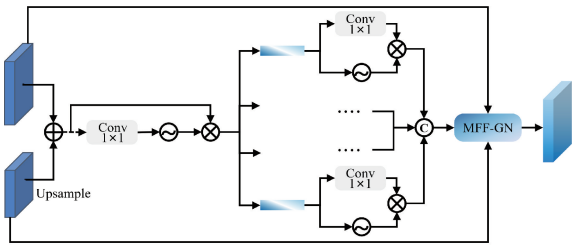


图 7　GFF 单元

Fig. 7　The unit of GFF

如图 8 所示，MFR 模块负责根据特征的重要性动态分配处理路径，实现特征的自适应重建。该模块首先通过全局自适应平均池化提取特征的全局统计信息，然后基于阈值比较策略将特征分为高权重路径和低权重路径进行差异化处理。对于全局特征 $\boldsymbol{X}_{\mathrm{global}}$，MFR 模块的重建过程公式表示为：

$$\boldsymbol{X}_{\mathrm{recon}}=\mathcal{F}_{\mathrm{high}}(\boldsymbol{X}_{\mathrm{global}}\odot\mathbb{I}_{w\geqslant\tau})+\mathcal{F}_{\mathrm{low}}(\boldsymbol{X}_{\mathrm{global}}\odot\mathbb{I}_{w<\tau})+\mathcal{F}_{\mathrm{norm}}(\boldsymbol{X}_{\mathrm{grouped}})\tag{11}$$

其中，$w=\sigma(\mathrm{GAP}(\boldsymbol{X}_{\mathrm{global}}))$ 表示通过全局平均池化和 Sigmoid 激活函数得到的权重系数，τ 为自适应阈值，$\mathbb{I}$ 为指示函数，$\mathcal{F}_{high}$ 和 $\mathcal{F}_{low}$ 分别表示高权重路径的逐点卷积操作和低权重路径的深度可分离卷积操作，$\mathcal{F}_{norm}$ 表示组归一化处理函数。

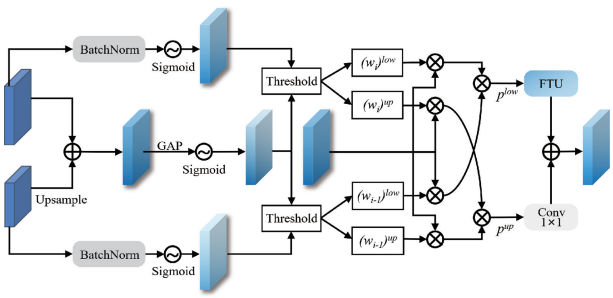


图 8　MFR 模块

Fig. 8　The module of MFR

EFC 特征融合策略通过引入分组特征聚焦单元和多级特征重建模块，构建了一个高效的自适应特征融合框架，解决了传统特征融合方法中存在的语义信息丢失和冗余特征干扰问题。相较于传统的 Concat 操作，EFC 策略能够更好地捕获小目标的细粒度特征，提升模型对航拍场景中复杂背景下小目标的检测性能。

1.5　IMIoU 损失函数

RT-DETR 所用的边界框损失函数（generalized iou，GIOU）在处理航拍无人机小目标检测任务时，对于 IoU 数值变化表现出较低的敏感性，这种敏感性不足导致模型在训练过程中难以精确捕捉小目标的边界信息变化。为改善这一关键问题，提出改进边界框损失函数（inner-modified penalty distance IoU，IMIoU），通过引入内部区域感知机制和多点距离惩罚策略，显著提升了模型对小目标边界预测的敏感性和精确性，从而有效改善了小目标检测的整体性能。

IMIoU 损失函数融合了 InnerIoU[29] 和 MPDIoU[30] 两种互补的几何约束机制。InnerIoU 通过构建缩放的内部重叠区域来增强模型对目标核心区域的感知能力，而 MPDIoU 则通过多点距离度量来约束预测框与真实框之间的几何关系。这种双重约束策略既保证了小目标检测的鲁棒性，又维持了边界框回归的精确性，为航拍无人机小目标检测提供了更加稳定且高效的优化目标。损失函数定义如图 9 所示。

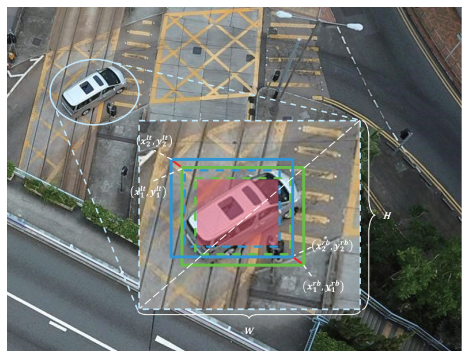


图 9　IMIoU 定义图

Fig. 9　The definition diagram of IMIoU

InnerIoU 通过引入缩放因子 *ratio* 对原始边界框进行内部收缩，构建更加紧密的核心区域来计算交并比，从而减少边界不确定性对损失函数的影响。具体而言，对于输入的边界框坐标，首先根据缩放因子计算内部边界框的坐标范围，然后基于收缩后的边界框计算交集和并集面积。InnerIoU 的基本计算公式定义为：

$$\mathrm{InnerIoU}=\frac{inter}{union}\tag{12}$$

内部交集面积的计算公式为：

$$inter=\max\begin{pmatrix}0,\min(x_1^{\mathrm{inner}},x_2^{\mathrm{inner}})\\-\max(x_1^{\mathrm{inner}},x_2^{\mathrm{inner}})\end{pmatrix}\times\max\begin{pmatrix}0,\min(y_1^{\mathrm{inner}},y_2^{\mathrm{inner}})\\-\max(y_1^{\mathrm{inner}},y_2^{\mathrm{inner}})\end{pmatrix}\tag{13}$$

内部并集面积的计算公式为:

$$union = w_1 \times h_1 \times ratio^2 + w_2 \times h_2 \times ratio^2 - inter + \varepsilon \tag{14}$$

其中,w_i 和 h_i 分别表示第 i 个边界框的宽度和高度,$ratio$ 为内部缩放因子,ε 为防止除零的常数。

MPDIoU 通过计算预测框与真实框对应顶点之间的欧几里得距离来引入几何惩罚项,确保边界框在空间位置上的精确对齐。该机制重点关注边界框左上角和右下角顶点的位置偏差,通过距离惩罚来约束预测框的几何形状和位置精度。MPDIoU 的惩罚项计算公式定义为:

$$\mathrm{MPD}_{penalty} = \frac{d_1}{h_w} + \frac{d_2}{h_w} = \frac{(x_2^{lt} - x_1^{lt})^2 + (y_2^{lt} - y_1^{lt})^2}{h_w} + \frac{(x_2^{rb} - x_1^{rb})^2 + (y_2^{rb} - y_1^{rb})^2}{h_w} \tag{15}$$

其中,d_1 和 d_2 分别表示左上角和右下角顶点之间的欧几里得距离平方,h_w 为归一化因子,下标 lt 和 rb 分别代表左上角和右下角坐标。

综合 InnerIoU 和 MPDIoU 的优势,IMIoU 损失函数的最终表达式定义为:

$$\mathcal{L}_{\mathrm{IMIoU}} = 1 - \mathrm{InnerIoU} + \frac{d_1^2 + d_2^2}{h_w},\ h_w = 2 \tag{16}$$

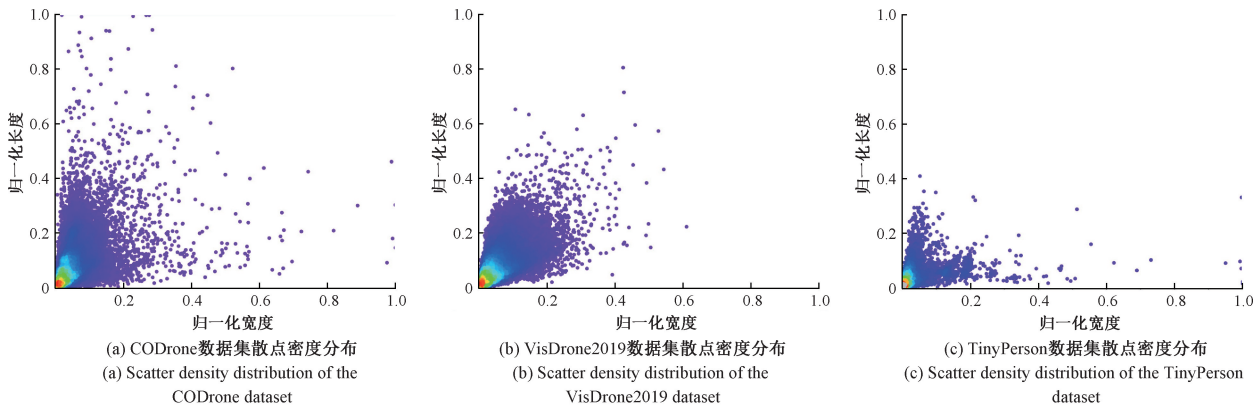
IMIoU 损失函数通过 InnerIoU 提供稳定的核心区域约束,同时借助 MPDIoU 的多点距离惩罚机制确保边界框的精确定位,形成了一个既鲁棒又精确的优化目标。该函数能够有效缓解小目标检测中的边界框敏感性问题,提高模型训练的稳定性,并在保持计算效率的同时显著提升航拍无人机场景下小目标的检测精度和定位准确性。

2 实验及结果分析

2.1 数据集

本实验共采用 3 个开源小目标数据集:VisDrone2019、CODrone 和 TinyPerson。数据集的散点密度分布分别如图 10(a)~(c)所示,数据集示例分别如图 10(d)~(f)所示。其中,VisDrone2019 数据集是专为无人机航拍图像目标检测设计,由天津大学 AISkyeYE 团队发布。该数据集包含 6 471 张用于训练的图像、1 610 张测试图像和 548 张验证图像,数据集涵盖了 10 种目标类别,如行人、汽车、面包车等,目标尺寸涵盖范围大且多数在复杂的背景环境以及不同的天气条件。鉴于无人机图像中的目标通常面临小尺度、遮挡和不均匀样本分布等挑战,VisDrone2019 数据集成为评估目标检测算法在复杂环境下表现的一个重要基准,对于验证小目标检测模型精度和鲁棒性具有现实意义。

CODrone 数据集是专为无人机航拍图像有向目标检测设计,由厦门大学团队发布。该数据集涵盖 12 种目标类别,数据集按照官方 5∶2∶3的比例划分为训练集、验证集和测试集,分别包含 5 002、2 001 和 3 001 张图像。数据采集覆盖 3 种飞行高度和两种相机角度的组合,在白天和夜间不同光照条件下拍摄,为评估算法在无人机航拍场景下的泛化性能提供了重要的测试基准。



(a) CODrone数据集散点密度分布
(a) Scatter density distribution of the CODrone dataset

(b) VisDrone2019数据集散点密度分布
(b) Scatter density distribution of the VisDrone2019 dataset

(c) TinyPerson数据集散点密度分布
(c) Scatter density distribution of the TinyPerson dataset

(d) CODrone 数据集示例图
(d) CODrone dataset example diagram

(e) VisDrone2019 数据集示例图
(e) VisDrone2019 dataset example diagram

(f) TinyPerson数据集示例图
(f) TinyPerson dataset example diagram

图 10 数据集散点密度分布及示例图

Fig. 10 The scatter density distribution and example diagram of the datasets

TinyPerson 数据集是专为小目标检测设计，由中科院大学团队发布。该数据集包含 1 610 张图像，其中 794 张图片作为训练集，816 张图片作为测试集。数据采集主要覆盖海洋和海滩场景，为海上搜救、远距离监控等应用提供了重要的评估基准，突出了小目标检测在复杂背景下的技术挑战。

2.2　实验环境与参数配置

本实验环境在 Ubuntu 22.04 操作系统上进行，硬件平台包括 AMD EPYC 9654 处理器、NVIDIA GeForce RTX 4090(24 GB 显存)显卡、40 GB 系统内存。软件环境方面，编译和运行环境 Python 3.10，深度学习框架 PyTorch 2.1.0，CUDA 版本为 12.1。在模型训练过程中，训练轮次设定为 300，批量大小为 8，数据加载器的 num_workers 设置为 8。优化器采用 AdamW，初始学习率设为 0.000 1，权重衰减默认为 0.000 1。

2.3　评价指标

本文采用了参数量(Params)、浮点运算量(GFLOPs)、准确率(Precision，P)、召回率(Recall，R)、平均精度均值(mean average precision，mAP)和每秒帧数(frames per second，FPS)作为模型的评价指标。

P 是指模型预测为正样本的样本中实际为正样本的比例；R 是指模型正确预测为正样本的样本占实际正样本的比例，计算公式为：

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

式中：TP 表示真阳性样本的数量，FP 表示假阳性样本的数量，FN 表示假阴性样本的数量。高准确率意味着模型对正样本的预测更加可靠。

平均精度均值是多类别平均精度(average precision，AP)的平均值，用于评估模型在所有类别上的整体性能。AP 是通过计算精确率-召回率(PR)曲线下的面积得到的，计算公式为：

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{19}$$

式中：$P(R)$ 表示在给定召回率 R 时的精确率。

mAP 计算公式为：

$$mAP = \frac{1}{K}\sum_{i=1}^{K} AP_i \tag{20}$$

式中：K 为类别数量；AP_i 表示第 i 个类别的平均精度。本文计算 mAP@0.5 和 mAP@0.5:0.95 分别表示在 IoU 阈值为 0.5 时的平均精度和在 IoU 从 0.5～0.95 的平均精度。

2.4　主干网络对比实验

为验证本文提出的 MCGA 主干网络的有效性，以 ResNet-18 作为基线模型，选取了近年来具有代表性的轻量化主干网络进行对比实验，包括 MobileNetV4-Conv-M、StarNet-S3、EfficientViT-M3、Fasternet-T1 和 RepViT-M1 等先进网络架构。实验结果如表 1 所示。从实验结果可以看出，各模型在不同评价指标上表现存在明显差异，在精确率上，本文提出的 MCGA 相比基线模型提升了 1.4%，同时优于其他所有对比网络。从召回率的指标分析得出 MCGA 的表现最为突出，相比基线模型提升了 1.6%，显著超越了 MobileNetV4-Conv-M 和 Starnet-S3。在目标检测的核心指标 mAP@0.5 当中，MCGA 达到 48.9%的最佳成绩，相比基线模型提升了 1.5%，比表现次优的 Fasternet-T1 高出 2.4%。最后，在 mAP@0.5:0.95 指标上，MCGA 相比基线模型提升了 1.1%，充分证明了其在不同 IoU 阈值下的检测精度优势。从模型复杂度角度分析，MCGA 的参数量相比基线模型减少了 27.1%，计算量虽然略高于部分轻量化网络，但考虑到其显著的精度提升，仍保持了良好的效率平衡。从检测速度角度分析，MCGA 的 FPS 为 74.3，虽然相比基线模型略有下降，这是由于引入门控聚合机制带来的额外计算开销，但其检测速度仍显著优于 EfficientViT-M3 和 RepViT-M1。实验结果表明了 MCGA 在保持模型轻量化的同时，在精确率、召回率和平均精度等关键指标上均取得了最优表现，实现了检测精度与模型复杂度的有效平衡，有效验证了所提出的主干网络改进策略对于航拍无人机小目标检测任务的有效性。实验结果表明了 MCGA 在保持模型轻量化的同时，在精确率、召回率和平均精度等关键指标上均取得了最优表现，有效验证了所提出的主干网络改进策略对于航拍无人机小目标检测任务的有效性。

表 1　不同主干网络对比结果

Table 1　The comparison results of different backbone networks

主干网络	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	Params/10^6	FPS(bs=1)
ResNet-18	60.7	46.3	47.4	28.8	57.0	19.9	**94.9**
MobileNetv4-Conv-M	59.5	42.3	43.1	25.9	39.5	**11.3**	84.1
Starnet-S3	59.3	42.3	43.6	26.4	**35.3**	14.1	77.4
Fasternet-T1	61.9	45.3	46.5	28.0	37.3	14.4	84.2
RepViT-M1	60.1	45.1	46.1	28.1	49.5	15.0	66.1
EfficientViT-M3	59.5	44.8	45.7	27.8	41.2	15.2	45.9
MCGA	**62.1**	**47.9**	**48.9**	**29.9**	37.9	14.5	74.3

2.5 损失函数对比实验

为验证本文提出的IMIoU损失函数的有效性,在基线模型的基础上使用VisDrone2019数据集进行多种IoU损失函数的对比实验,实验结果如表2所示。基线模型采用GIoU作为默认的IoU损失函数,为全面评估IMIoU的性能表现,选取了目前广泛应用的IoU、CIoU、ShapeIoU、InnerIoU和MPDIoU损失函数进行对比分析。实验结果表明,本文提出的IMIoU损失函数在两个关键指标上均取得了最优性能,mAP@0.5达到48.2%,相比基线GIoU的47.4%提升了0.8%,mAP@0.5:0.95达到29.8%,相比基线的28.8%提升了1.0%。相较于传统IoU损失函数,IMIoU能够更加精确地感知目标边界变化并实现更准确的边界框定位,显著改善了小目标的边界识别效果。这充分证明了IMIoU通过融合InnerIoU内部区域感知和MPDIoU多点距离约束机制能够有效提升小目标检测的精度和鲁棒性,验证了该损失函数在航拍无人机小目标检测任务中的优越性和实用性。

表2 不同IoU对比结果

Table 2 The comparison results of different IoU

IoU	GIoU	CIoU	ShapeIoU	InnerIoU	MPDIoU	IMIoU
mAP@0.5/%	47.4	47.5	47.6	46.7	47.6	48.2
mAP@0.5:0.95/%	28.8	28.7	28.9	28.1	29.0	29.8

2.6 消融实验

本文采用了多种方法对RT-DETR模型进行改进,为验证本文提出的各改进模块的有效性,在公开的VisDrone2019小目标数据集上进行了消融实验。实验结果如表3所示。基准RT-DETR模型在该数据集上的mAP@0.5为47.4%,mAP@0.5:0.95为28.8%。通过逐步添加改进模块进行消融分析发现,MCGA模块能够有效提升检测精度,使mAP@0.5提升1.5%,mAP@0.5:0.95提升1.1%,同时显著减少参数量27.1%和计算量13.2%,体现了轻量化设计的优势;Micro-OmniPyramid(MOP)模块相较于基线模型mAP@0.5提升2.0%,mAP@0.5:0.95提升1.5%;EFC模块使用相较基线模型mAP@0.5小幅提升0.1%,mAP@0.5:0.95提升0.4%;IMIoU损失函数的改进使mAP@0.5提升0.8%,mAP@0.5:0.95提升1.0%。

表3 消融实验结果

Table 3 The results of the ablation experiment

序号	MCGA	MOP	EFC	IMIoU	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Params/10^6	GFLOPs	FPS(bs=1)
1	×	×	×	×	60.7	46.3	47.4	28.8	19.9	57.0	**94.9**
2	√	×	×	×	62.1	47.9	48.9	29.9	**14.5**	**49.5**	74.3
3	×	√	×	×	62.7	48.1	49.4	30.3	20.5	65.2	86.7
4	×	×	√	×	61.4	45.9	47.5	29.2	20.8	63.6	71.9
5	×	×	×	√	61.9	46.4	48.2	29.8	19.9	57.0	92.8
6	√	√	×	×	63.0	49.6	50.5	31.2	15.7	65.2	69.6
7	√	√	√	×	63.9	49.7	50.9	31.6	17.2	68.2	59.6
8	√	√	√	√	**63.6**	**49.8**	**51.3**	**31.9**	17.2	68.2	60.5

最终改进的MGEF-DETR模型相比于基线模型在检测精度上实现了显著提升,mAP@0.5提升3.9%,mAP@0.5:0.95提升3.1%,同时精确度提升2.9%,召回率提升3.5%,检测速度达到60.5 fps,在精度显著提升的同时保持了良好的实时检测性能,能够有效解决航拍场景下无人机小目标检测精度不足、特征表达能力弱、多尺度目标融合效果差等问题。

2.7 不同模型对比实验

为进一步验证改进算法MGEF-DETR模型在航拍小目标检测任务中的先进性,本研究在VisDrone2019、TinyPerson和CODrone 3个数据集上进行了充分的对比实验,实验结果分别如表4~6所示。在VisDrone2019数据集上,选取了涵盖单阶段、双阶段以及基于Transformer架构的多种主流检测算法进行对比,包括YOLO系列、RetinaNet、TOOD、GFL、YOLOX等单阶段检测器,Faster R-CNN和Cascade R-CNN等双阶段检测器,DDQ-DETR[31]、DINO[32]等Transformer模型变体,以及DMU-YOLO[33]和MSM-DETR[34]等主流目标检测模型算法。在TinyPerson和CODrone数据集上,主要选取了YOLO系列最新模型(YOLOv8m至YOLOv12m)以及基于Transformer架构的DINO模型进行对比验证。

在VisDrone2019数据集上,MGEF-DETR相较于基线模型RT-DETR在各项指标上均有显著提升,体现了改进方法在性能提升和模型轻量化方面的双重优势。与其

表 4　VisDrone2019 数据集对比实验结果

Table 4　The comparison experiment results of the VisDrone2019 dataset

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Params/10^6	GFLOPs	FPS(bs=1)
Faster R-CNN	50.5	36.6	39.3	23.3	41.4	208.0	60.2
Cascade R-CNN	50.7	36.5	39.6	23.9	69.3	236.0	48.4
RetinaNet	47.7	38.0	37.0	22.2	36.5	210.0	62.5
TOOD	47.1	38.0	36.7	22.0	32.0	199.0	47.7
DDQ-DETR[31]	53.8	45.3	44.7	26.0	—	—	22.0
DINO[32]	49.7	43.4	41.8	23.6	47.6	274.0	29.2
GFL	45.5	36.4	35.2	20.9	32.3	206.0	62.6
YOLOX	52.0	39.2	39.5	22.9	25.3	73.8	119.7
YOLOv5m	54.0	40.6	42.4	25.7	25.1	64.0	253.1
YOLOv5l	54.5	42.9	43.8	26.8	53.1	134.7	188.4
YOLOv8s	49.9	38.4	39.0	23.2	**11.1**	**28.5**	**336.0**
YOLOv8m	53.8	41.3	42.2	25.7	25.8	78.7	257.8
YOLOv9m	55.1	42.6	43.9	26.8	20.0	76.5	170.5
YOLOv9c	54.5	43.4	44.7	27.1	25.3	102.4	153.3
YOLOv10m	54.7	41.1	42.7	26.0	15.3	58.9	213.2
YOLOv10b	54.8	42.1	43.9	27.0	19.0	91.7	209.2
YOLOv10l	56.4	43.0	45.1	27.8	24.3	120.0	175.0
YOLOv11m	54.5	43.2	44.4	27.0	20.0	67.7	208.4
YOLOv11l	55.8	43.1	44.6	27.5	25.3	86.6	137.0
YOLOv12m	53.7	42.0	43.4	26.3	20.1	67.2	163.1
YOLOv12l	56.1	43.0	44.7	27.6	26.3	88.6	99.3
RT-DETR	60.7	46.3	47.4	28.8	19.9	57.0	94.9
文献[17]	63.2	47.3	49.4	30.2	33.8	—	54.1
EMRT-DETR[18]	62.0	46.6	48.8	30.5	13.8	68.1	52.0
DMU-YOLO[33]	58.6	47.4	48.1	29.7	25.4	89.5	—
MSM-DETR[34]	—	—	49.5	30.6	22.2	72.9	81.0
MGEF-DETR(本文)	**63.6**	**49.8**	**51.3**	**31.9**	17.2	68.2	60.5

表 5　TinyPerson 数据集对比实验结果

Table 5　The comparison experiment results of the TinyPerson dataset

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Params/10^6	GFLOPs	FPS(bs=1)
YOLOv8m	36.9	18.8	17.4	6.2	25.8	78.7	**260.1**
YOLOv9m	38.4	18.5	17.4	6.2	20.0	76.5	171.7
YOLOv10b	37.1	19.1	17.6	6.3	19.0	91.6	211.2
YOLOv11m	38.4	19.4	18.0	6.6	20.0	67.7	212.9
YOLOv12m	38.5	20.0	19.0	6.9	20.1	67.1	169.0
DINO[32]	16.6	19.7	8.6	2.8	47.6	274.0	27.5
RT-DETR	40.1	24.0	19.8	7.0	19.9	**56.9**	94.5
MGEF-DETR(本文)	**42.5**	**25.5**	**21.5**	**7.7**	**17.1**	68.2	60.4

他主流模型相比，在精确率方面，MGEF-DETR 超越了 YOLO 系列最高值(YOLOv10l)7.2%。在 mAP@0.5 指标上，MGEF-DETR 取得了 51.3%的最高性能，超越了同样基于 Transformer 架构的 MSM-DETR、YOLO 系列表现最佳的 YOLOv10l，相比双阶段算法中表现最好的 Cascade R-CNN 提升了 11.7%，证明了改进模型在目标定位精度方面的显著优势。在 mAP@0.5:0.95 指标上，MGEF-DETR 大幅超越了 Faster R-CNN、Cascade R-CNN

表 6 CODrone 数据集对比实验结果

Table 6 The comparison experiment results of the CODrone dataset

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Params/10^6	GFLOPs	FPS(bs=1)
YOLOv8m	41.0	29.7	27.4	14.0	25.8	78.7	**258.5**
YOLOv9m	42.8	30.4	28.1	14.4	20.0	76.6	168.6
YOLOv10b	42.9	30.2	28.7	14.9	19.0	91.7	205.0
YOLOv11m	42.7	31.3	29.4	15.1	20.0	67.7	207.4
YOLOv12m	42.0	30.8	28.8	14.8	20.1	67.2	161.2
DINO[32]	38.4	32.7	30.5	15.3	47.6	274.0	28.1
RT-DETR	44.5	34.5	31.5	16.0	19.9	**57.0**	93.9
MGEF-DETR(本文)	**43.7**	**36.5**	**33.9**	**17.6**	**17.2**	68.2	61.0

和 DINO 等主流检测器,提升幅度超过 8%。与同样基于 Transformer 架构的 MSM-DETR 相比,MGEF-DETR 在 mAP@0.5:0.95 指标上提升 1.3%的同时,参数量减少 22.5%,检测速度达到 60.5 fps,显著优于 DINO 和 DDQ-DETR,在精度与速度上实现了良好平衡。

在 TinyPerson 数据集上,MGEF-DETR 相较于基线模型 RT-DETR 精确率提升 2.4%,召回率提升 1.5%,mAP@0.5 提升 1.7%,mAP@0.5:0.95 提升 0.7%。在与其他模型的对比中,MGEF-DETR 展现出突出的小目标检测能力。相较于基于 DINO 模型,MGEF-DETR 在 mAP@0.5 指标上实现了 12.9%的大幅提升,充分说明了改进算法在处理极小目标时的优越性。与 YOLO 系列最新模型 YOLOv12m 相比,MGEF-DETR 在各项指标上均有显著提升,尤其在召回率方面提升了 5.5%,同时检测速度为 60.4 fps,相比 DINO 提升明显,验证了改进模型的实用性。

在 CODrone 数据集上,MGEF-DETR 相较于基线模型 RT-DETR 在目标捕获能力方面表现尤为突出,召回率提升 2.0%,mAP@0.5 提升 2.4%,mAP@0.5:0.95 提升 1.6%。与其他模型的对比结果显示 MGEF-DETR 在复杂无人机检测场景下具有显著优势,与 DINO 模型相比,MGEF-DETR 在 mAP@0.5:0.95 指标上提升 2.3%,同时参数量仅为 DINO 的 36.1%,检测速度是 DINO 的两倍以上,充分展现了改进模型的架构优势。

综合 3 个数据集的对比实验结果表明,MGEF-DETR 在不同类型的航拍小目标检测任务中均展现出了优异的性能。相较于基线模型 RT-DETR,改进模型在保持轻量化优势的同时,在所有关键性能指标上均实现了显著提升,验证了所提出改进方法的有效性。与当前主流的单阶段、双阶段以及基于 Transformer 的检测算法相比,MGEF-DETR 均展现出显著的性能优势,尤其在小目标检测精度和模型效率方面取得了良好的平衡,充分证明了改进算法在航拍复杂场景下的鲁棒性和先进性。

2.8 特征融合特征图分析

为深入分析 EFC 特征融合模块相较于传统 Concat 操作在特征表达和信息整合方面的优势,通过特征图可视化技术对两种融合策略的效果进行对比分析。特征图可视化结果如图 11 所示,分别展示了在 VisDrone2019 如图 11(a)所示、CODrone 如图 11(b)所示和 TinyPerson 如图 11(c)所示 3 个数据集上 Concat 操作与 EFC 模块的特征激活模式差异。

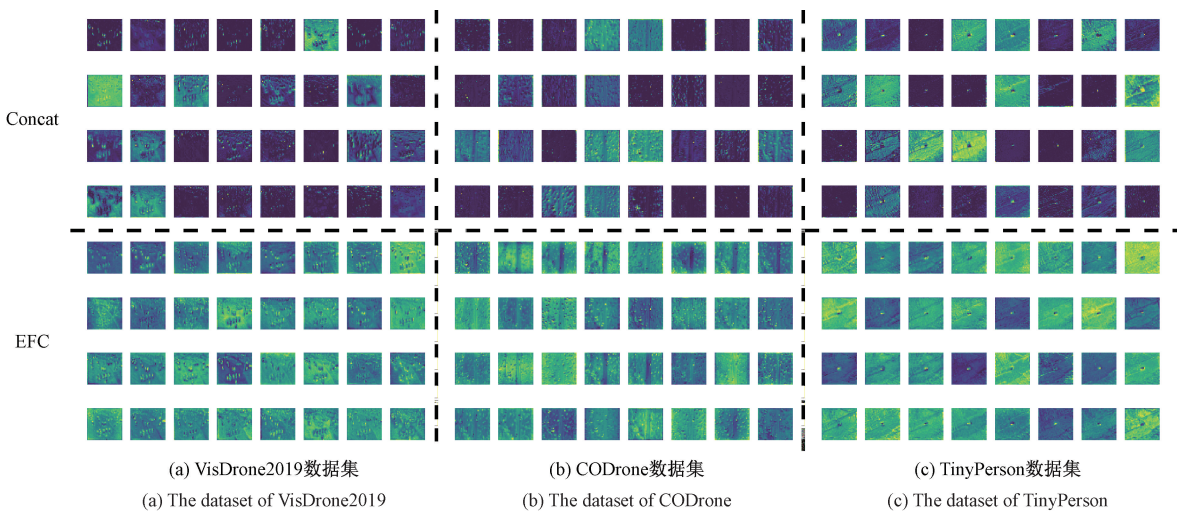


图 11 不同特征融合方法的特征映射可视化

Fig. 11 The feature map visualization of different feature fusion methods

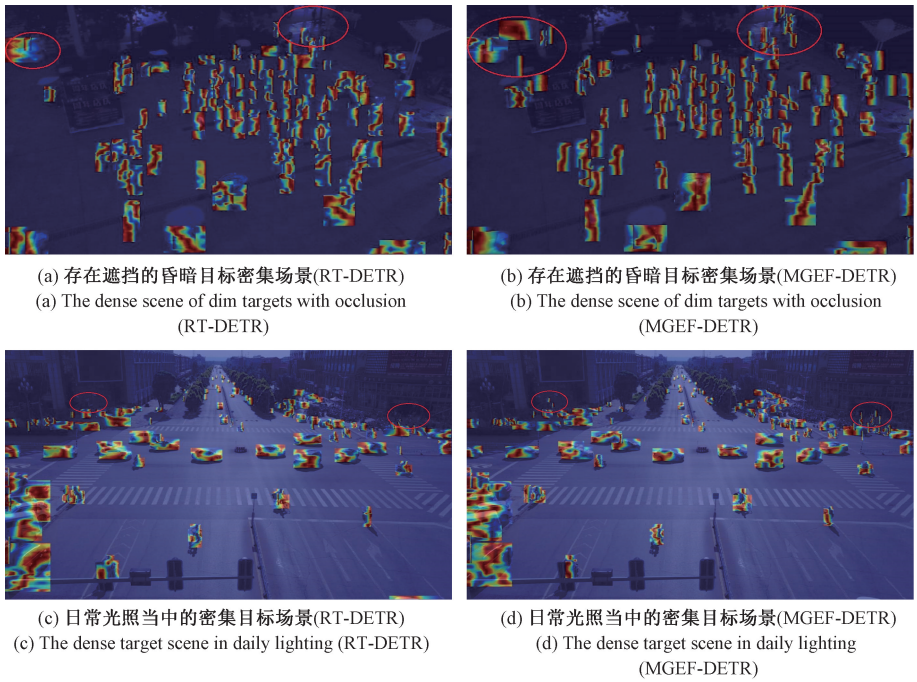
从可视化结果可以明显观察到，传统 Concat 操作产生的特征图在小目标区域表现出较为分散和不连续的激活模式，特别是在图像中那些尺寸微小的目标区域（以绿色和蓝绿色高亮显示），激活响应呈现零散分布，缺乏有效的空间连贯性。这种现象表明简单的特征拼接操作难以充分建模不同尺度特征间的语义关联，导致小目标的关键特征信息在融合过程中出现弱化甚至丢失。

相比之下，EFC 特征融合模块在相同的小目标区域展现出显著改善的特征表达能力。可视化结果显示，EFC 产生的特征图在小目标位置呈现更加集中、连贯且强烈的激活响应，特征激活区域不仅在空间上更加紧密，而且在强度上表现出更高的响应值。这种改善得益于 EFC 模块中 GFF 和 MFR 的协同作用：GFF 通过多层次分组注意力机制有效捕获异构语义层次的特征激活模式，而 MFR 则通过自适应权重分配动态调控多尺度特征的贡献权重，从而实现更加精准和判别性的特征融合效果。

在 Micro-OmniPyramid 架构设计中，P2 层与 P3 层的融合仍采用传统 Concat 操作而非 EFC 模块。由于 EFC 模块设计为双输入架构，专门针对两个不同尺度特征图间的自适应融合进行优化，而 P2 与 P3 层的融合属于相邻尺度间的直接信息传递，更适合采用简单高效的拼接操作；P2 层包含丰富的高分辨率细节信息，通过 SPD 卷积稀疏编码策略已经实现了有效的特征压缩和表征，此时引入复杂的融合机制可能带来不必要的计算开销而效果提升有限；最后，实验表明在 P2-P3 融合阶段使用 Concat 操作能够保持特征传递的稳定性，为 EFC 模块在更深层特征融合中发挥作用奠定良好基础。

2.9　热力图分析

为深入评估模型的检测性能，引入 GradCAM++[35] 可视化技术进行定性分析，通过热力图直观展示网络对图像中不同区域的注意力分布情况，从而增强算法的可解释性，热力图结果如图 12 所示。首先，在昏暗目标密集且存在遮挡的复杂场景中，如图 12(a)所示，RT-DETR 热力图显示其对图像上方被遮挡行人的关注度较低，导致该区域出现明显的漏检现象，模型难以有效识别被部分遮挡的小目标；而图 12(b)所示的 MGEF-DETR 热力图则表现出对该被遮挡行人区域显著增强的关注度，成功解决了漏检问题并准确感知到被遮挡的行人目标。其次，在日常光照条件下的密集目标场景中，图 12(c)所示的 RT-DETR 热力图显示其对位于图像左侧和右侧边缘区域的行人关注度不足，出现了边缘小目标的漏检情况；相比之下，图 12(d)所示的 MGEF-DETR 热力图在这些边缘区域表现出更强的激活响应和更集中的注意力分布，准确检测到了原本被漏检的行人目标。MGEF-DETR 通过增强特征提取和信息融合能力，显著提升了对复杂场景中小目标的感知能力和检测精度，验证了所提出改进方法的有效性。



(a) 存在遮挡的昏暗目标密集场景(RT-DETR)
(a) The dense scene of dim targets with occlusion (RT-DETR)

(b) 存在遮挡的昏暗目标密集场景(MGEF-DETR)
(b) The dense scene of dim targets with occlusion (MGEF-DETR)

(c) 日常光照当中的密集目标场景(RT-DETR)
(c) The dense target scene in daily lighting (RT-DETR)

(d) 日常光照当中的密集目标场景(MGEF-DETR)
(d) The dense target scene in daily lighting (MGEF-DETR)

图 12　热力图对比结果

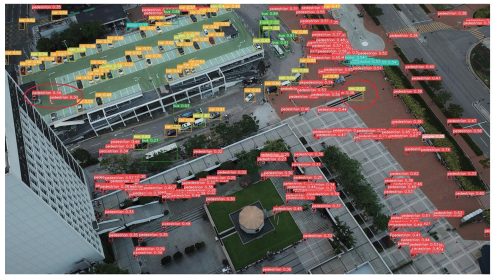
Fig. 12　The comparison results of the heatmaps

2.10　检测结果分析

从检测结果图的检测效果可以直观地看出改进模型相较于原始 RT-DETR 的显著提升效果，检测效果如图 13 所示。在日常光照条件下的密集目标场景中，图 13(a)所示的 RT-DETR 基线模型检测结果存在将道路障碍物误识别为行人、摩托车错误分类为三轮车的误检现象以及漏

(a) 日常光照当中的密集目标场景(RT-DETR)
(a) The dense target scene in daily lighting (RT-DETR)

(b) 日常光照当中的密集目标场景(MGEF-DETR)
(b) The dense target scene in daily lighting (MGEF-DETR)

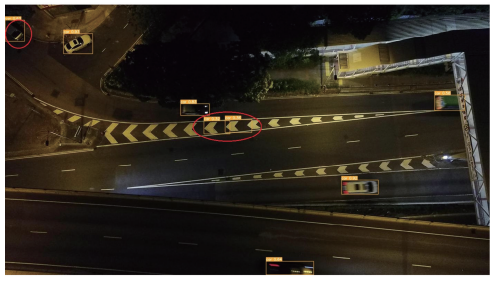


(c) 日常光照当中的高空俯视密集目标场景(RT-DETR)
(c) The dense target scene of high-altitude overhead view in daily lighting (RT-DETR)

(d) 日常光照当中的高空俯视密集目标场景(MGEF-DETR)
(d) The dense target scene of high-altitude overhead view in daily lighting (MGEF-DETR)

(e) 夜间复杂光照当中的高速路段场景(RT-DETR)
(e) The highway section scene of complex lighting at night (RT-DETR)

(f) 夜间复杂光照当中的高速路段场景(MGEF-DETR)
(f) The highway section scene of complex lighting at night (MGEF-DETR)



(g) 夜间复杂光照当中的高空俯视高速路段场景(RT-DETR)
(g) The highway section scene of high-altitude overhead view under complex lighting at night (RT-DETR)

(h) 夜间复杂光照当中的高空俯视高速路段场景(MGEF-DETR)
(h) The highway section scene of high-altitude overhead view under complex lighting at night (MGEF-DETR)

图 13 不同场景下检测结果对比图

Fig. 13 The comparison diagram of detection results under different scenarios

检多辆带篷三轮车;而图 13(b)所示的 MGEF-DETR 模型检测结果有效避免了将路障误识别为行人的情况,同时成功准确识别出摩托车类别而非三轮车,并且准确的识别出了多辆带篷三轮车。在日常光照条件下的高空俯视密集目标场景中,图 13(c)所示的 RT-DETR 基线模型检测图像中出现了将地面标识、建筑物屋顶等非人体目标误识别为行人的误检问题,同时存在将道路下水道误分类为面包车的错误检测结果;相比之下,图 13(d)所示的 MGEF-DETR 模型检测结果有效避免了上述误检问题,准确区分了不同类别的目标对象。在夜间复杂光照条件下的高速路段场景中,图 13(e)所示的 RT-DETR 基线模型将位于图像左下角区域的路灯设施误检测为卡车目标,反映出基

线模型在低光照环境下的特征提取能力不足；而图 13(f)所示的 MGEF-DETR 模型成功避免了将路灯误检测为卡车的情况，保持了较高的检测准确性。此外，图 13(g)所示的 RT-DETR 基线模型检测结果中存在将位于图像左上方的面包车误分类为小汽车以及将地面交通标识误检测为小汽车的问题；图 13(h)所示的 MGEF-DETR 模型则准确识别出左上方车辆为面包车类别，且未出现将地面标识误检测为车辆的情况，体现出更强的目标分类精度。综合上述检测场景，MGEF-DETR 模型在复杂场景下对小目标的识别准确性得到显著提升，有效解决了基线模型在目标检测和分类任务中存在的误检和误分类问题。

3　结　　论

航拍无人机场景中存在的高空俯视视角、复杂背景干扰、多尺度目标混合以及光照条件多变等特殊环境因素对小目标检测的精度和实时性存在挑战，现有 RT-DETR 模型在此类应用中出现检测精度不足和特征表达能力有限等瓶颈。为此，本文提出改进模型 MGEF-DETR。本研究从网络架构设计和优化策略两个维度进行系统性改进，MCGA 主干网络有效提升了模型对小目标细节特征的感知能力；Micro-OmniPyramid 金字塔网络强化了多尺度特征的有效整合；EFC 模块优化了特征间的语义关联性表达；IMIoU 损失函数增强了边界框回归的准确性。通过对比实验和消融实验验证了各改进模块的有效性，MGEF-DETR 在主流数据集上实现了检测精度的显著提升，在与当前先进检测算法的比较中展现出明显的性能优势。跨数据集泛化实验进一步证实了模型的稳定性和适用性，表明所提方法具备良好的实际应用潜力。尽管如此，当前模型对严重遮挡场景的处理能力仍有待进一步增强。未来研究将重点关注自适应特征学习和多任务联合优化机制，以提升模型在实际航拍应用中的部署效果和检测性能。

参考文献

[1] ZOU ZH X, CHEN K Y, SHI ZH W, et al. Object detection in 20 years: A survey[C]. IEEE, 2023, 111(3): 257-276.

[2] CHENG N, WU SH, WANG X, et al. AI for UAV-assisted IoT applications: A comprehensive review[J]. IEEE Internet of Things Journal, 2023, 10(16): 14438-14461.

[3] SRIVASTAVA S, DIVEKAR A V, ANILKUMAR C, et al. Comparative analysis of deep learning image detection algorithms[J]. Journal of Big Data, 2021, 8: 66.

[4] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.

[5] 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, 2016, 39(6): 1137-1149.

[6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. 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]. 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]. IEEE International Conference on Computer Vision, 2017: 2999-3007.

[9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.

[10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision, 2020: 213-229.

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

[12] 吴一全，童康. 基于深度学习的无人机航拍图像小目标检测研究进展[J]. 航空学报，2025，46(3)：181-207.
WU Y Q, TONG K. Research advances on deep learning-based small object detection in UAV aerial images[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(3): 181-207.

[13] 邓天民，程鑫鑫，刘金凤，等. 基于特征复用机制的航拍图像小目标检测算法[J]. 浙江大学学报(工学版)，2024，58(3)：437-448.
DENG T M, CHENG X X, LIU J F, et al. Small target detection algorithm for aerial images based on feature reuse mechanism[J]. Journal of Zhejiang University (Engineering Science), 2024, 58(3): 437-448.

[14] 谢椿辉，吴金明，徐怀宇. 改进 YOLOv5 的无人机影像小目标检测算法[J]. 计算机工程与应用，2023，59(9)：198-206.
XIE CH H, WU J M, XU H Y. Small object detection algorithm based on improved YOLOv5 in UAV image[J]. Computer Engineering and Applications, 2023, 59(9): 198-206.

[15] 刘洋，任旭虎，刘宝弟，等. 基于 LDF-YOLO 的小目标检测方法[J]. 电子测量技术，2025，48(12)：156-165.
LIU Y, REN X H, LIU B D, et al. Small object detection method based on LDF-YOLO[J]. Electronic Measurement Technology, 2025, 48(12): 156-165.

[16] 贺智轩，陈里里，王翔，等. DMF-YOLOv11：基于改进 YOLOv11n 的无人机航拍图像目标检测算法[J]. 计算机工程与应用，2025，61(14)：88-100.

HE ZH X, CHEN L L, WANG X, et al. DMF-YOLOv11: Target detection algorithm for UAV images based on improved YOLOv11n[J]. Computer Engineering and Applications, 2025, 61(14): 88-100.

[17] 胡佳乐,周敏,申飞.面向无人机小目标的RTDETR改进检测算法[J].计算机工程与应用,2024,60(20):198-206.
HU J L, ZHOU M, SHEN F. Improved detection algorithm of RTDETR for UAV small target[J]. Computer Engineering and Applications, 2024, 60(20): 198-206.

[18] 姜贸翔,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.
JIANG M X, SI ZH J, WANG X ZH. Improved target detection algorithm for UAV images with RT-DETR[J]. Computer Engineering and Applications, 2025, 61(1): 98-108.

[19] 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.

[20] 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 Iinternational Conference on Computer Vision Workshops, 2019:213-226.

[21] YE K, TANG H D, LIU B W, et al. More clear, more flexible, more precise: A comprehensive oriented object detection benchmark for UAV[J]. ArXiv preprint arXiv:2504.20032, 2025.

[22] YU X H, GONG Y Q, JIANG N, et al. Scale match for tiny person detection[C]. 2020 IEEE/CVF Winter Conference on Applications of Computer Vision, 2020: 1246-1254.

[23] LI S Y, WANG Z D, LIU Z CH, et al. Moganet: Multi-order gated aggregation network[J]. ArXiv preprint arXiv:2211.03295, 2022.

[24] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 1571-1580.

[25] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017:936-994.

[26] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects[C]. Machine Learning and Knowledge Discovery in Databases, 2022: 443-459.

[27] CUI Y N, REN W Q, KNOLL A. Omni-kernel modulation for universal image restoration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024,34(12):12496-12509.

[28] XIAO Y, XU T F, YU X, et al. A lightweight fusion strategy with enhanced inter-layer feature correlation for small object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024,62:1-11.

[29] 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.

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

[31] ZHANG SH L, WANG X J, WANG J Q, et al. Dense distinct query for end-to-end object detection[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7329-7338.

[32] ZHANG H, LI F, LIU SH L, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. ArXiv preprint arXiv:2203.03605, 2022.

[33] 韩佰轩,彭月平,郝鹤翔,等.DMU-YOLO:机载视觉的多类异常行为检测算法[J].计算机工程与应用,2025,61(7):128-140.
HAN B X, PENG Y P, HAO H X, et al. DMU-YOLO: Multi-class abnormal behavior detection algorithm based on air-borne vision[J]. Computer Engineering and Applications, 2025, 61(7): 128-140.

[34] 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 Y W, JIANG Y, WANG Q K, et al. Research on real-time Transformer for multi-scale feature optimization in drone aerial imaging[J]. Computer Engineering and Applications, 2025, 61(9): 221-229.

[35] CHATTOPADHAY A, SARKAR A, HOWLADER P, et al. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks[C]. 2018 IEEE Winter Conference on Applications of Computer Vision(WACV), 2018: 839-847.

作者简介

侯林杰,硕士研究生,主要研究方向为计算机视觉、目标检测。

E-mail:2024720774@yangtzeu.edu.cn

卢承方,硕士研究生,主要研究方向为深度学习、目标检测。

E-mail:2023710700@yangtzeu.edu.cn

崔艳荣(通信作者),博士,教授,硕士研究生导师,主要研究方向为人工智能、信息处理。

E-mail:cyanr@yangtzeu.edu.cn