◎目标检测专题◎

MBFE-DETR:多尺度边界特征增强下的无人机目标检测算法

张 晞[1,2],赖惠成[1,2+],姜 迪[1,2],汤静雯[1,2],高古学[3],袁婷婷[1,2],聂 源[1,2]
1.新疆大学 计算机科学与技术学院,乌鲁木齐 830046
2.新疆大学 新疆维吾尔自治区信号检测与处理重点实验室,乌鲁木齐 830046
3.淮阴工学院 计算机与软件工程学院,江苏 淮安 223001
+ 通信作者 E-mail:lai@xju.edu.cn

摘 要:针对无人机视角下背景复杂、小目标比例较高且样本不平衡等问题,提出一种基于改进RT-DETR的无人机目标检测算法MBFE-DETR。设计一种基于C2f和单头自注意力模块的轻量化主干网络,降低模型参数量的同时提升网络的特征提取能力。提出多尺度边界特征增强协同网络MBFECN,通过其特有的多尺度边界特征增强机制和高效特征融合策略,解决了原模型在保持小目标边界细节方面的不足。引入Focaler-MPDIoU考虑框的位置匹配关系,同时通过线性区间映射重构原有IoU损失,使模型在复杂场景下的定位效果更好。针对样本不平衡的问题,采用新的分类损失函数ESVLoss,对分类损失值进行分段加权调整,并结合指数移动平均机制对权重进行动态平滑更新,使模型更具适应性。实验结果表明,在VisDrone2019-DET和DOTAv1.0数据集上,MBFE-DETR算法的mAP_{50}分别提升3.9和2.9个百分点,同时参数量减少了21.6%。

关键词:无人机目标检测;RT-DETR;单头自注意力;边界特征增强
文献标志码:A **中图分类号:**TP391.41 doi:10.3778/j.issn.1002-8331.2503-0307

MBFE-DETR: Multi-Scale Boundary Feature Enhancement for Drone Target Detection Algorithm

ZHANG Xi[1,2], LAI Huicheng[1,2+], JIANG Di[1,2], TANG Jingwen[1,2], GAO Guxue[3], YUAN Tingting[1,2], NIE Yuan[1,2]
1.School of Computer Science and Technology, Xinjiang University, Urumqi 830046, China
2.The Key Laboratory of Signal Detection and Processing, Xinjiang Uygur Autonomous Region, Xinjiang University, Urumqi 830046, China
3.College of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an, Jiangsu 223001, China

Abstract: Aiming at problems such as complex backgrounds, high proportion of small targets, and sample imbalance in drone perspective views, an improved drone object detection algorithm based on RT-DETR called MBFE-DETR is proposed. Firstly, a lightweight backbone network based on C2f and single-head self-attention modules is designed, reducing model parameters while enhancing feature extraction capabilities. Secondly, a multi-scale boundary feature enhancement collaborative network (MBFECN) is proposed, which addresses the original model's deficiencies in preserving small target boundary details through its unique multi-scale boundary feature enhancement mechanism and efficient feature fusion strategy. Then, Focaler-MPDIoU is introduced to consider the positional matching relationship between bounding boxes, while reconstructing the original IoU loss through linear interval mapping, resulting in better localization performance in complex scenes. Finally, to address sample imbalance, a new classification loss function called ESVLoss is adopted, which applies segmented weighted adjustments to classification loss values and combines an exponential moving average mechanism to dynamically update weights smoothly, making the model more adaptive. Experimental results show that on the VisDrone2019-DET and DOTAv1.0 datasets, the MBFE-DETR algorithm improves mAP_{50} by 3.9 and 2.9 percentage points respectively, while reducing parameters by 21.6%.

Key words: UAV object detection; RT-DETR; single-head self-attention; boundary feature enhancement

基金项目:新疆维吾尔自治区重点研发计划(2022B01008)。

作者简介:张晞(2000—),男,硕士研究生,研究方向为图像处理、无人机目标检测;赖惠成(1963—),男,教授,研究方向为视频/图像信息处理、图像理解与识别。

收稿日期:2025-03-26 **修回日期:**2025-07-07 **文章编号:**1002-8331(2025)17-0089-13

随着无人机技术的迅猛发展,无人机已被广泛运用于航拍观光、灾害救援、军事侦察、农业监测等领域。然而,由于无人机拍摄视角不定和硬件限制等原因,拍摄图像往往存在背景复杂、小目标占比高、分辨率低,以及样本数据分布不平衡等问题。这些因素导致主流的目标检测模型容易发生漏检和误检。

目前主流的目标检测算法主要分为两类:一类是先生成候选框,再在这些候选框内识别物体;另一类则直接获取预测结果,无须中间的区域检测过程。前者以R-CNN[1]、Fast R-CNN[2]、Faster R-CNN[3]为代表,被称为两阶段方法,但由于生成大量边界框会降低检测效率,难以有效满足无人机目标检测的实时性要求。后者主要包括YOLO系列[4-9]、SSD[10]、EfficientDet[11]和RefineDet[12]等,被称为单阶段方法。相较于两阶段检测算法,单阶段检测算法具有较高的实时性,许多学者对此进行了研究。Sahin等人[13]将YOLOv3的输出层扩充至五个以提高航拍小目标检测精度,但会导致参数量过大。李成豪等人[14]通过设计多尺度感受野融合模块提升了小目标检测精度,但多个串联的小卷积核在某些场景下仍难以完全替代大卷积核。聂源等人[15]设计级联双向特征金字塔KBiFPN,以及多级感受野特征聚合模块MFA,提高了航拍小目标检测精度。汤静雯等人[16]通过高效特征提取模块EM,增加小目标检测头等,增强了网络对小目标的识别能力。Tang等人[17]通过轻量级共享检测头和选择性双向扩散金字塔网络实现了更高的精度和更低的参数量。Wang等人[18]提出一种基于YOLOv8的轻量级小目标检测算法LSOD-YOLO,采用轻量级上采样器Dysample以最小的计算成本保留了丰富的图像细节。李峻宇等人[19]在标准的YOLOv5模型基础上,引入通道-细节注意力模块汇聚红外小目标的通道信息和细节信息,提高回归精度。张浩晨等人[20]结合可变形卷积和SPDConv模块提取P2层高分辨率信息,捕获更精细的边缘特征。尽管上述改进有效解决了一些问题,但大多数都依赖非极大值抑制技术来处理冗余的边界框,要选择合适的NMS阈值以适应不同场景,从而避免丢失目标的问题,效率有待进一步提升,并且CNN在捕捉长距离依赖信息时也存在一定局限性。

为进一步克服上述局限,利用Transformer[21]在全局上下文信息捕捉方面的优势来应对目标检测任务成为新趋势。DETR(detection Transformer)[22]取消了anchor机制和NMS处理,通过Transformer的全局特性对图像直接进行集合预测,简化了目标检测流程,但其参数量大且训练耗时较长。Zhu等人[23]提出了Deformable DETR,引入可变形注意力以降低了DETR的计算成本。Li等人[24]提出了DN-DETR,利用查询降噪机制改善匈牙利匹配二义性带来的收敛速度慢的问题。百度提出了实时目标检测Transformer(real-time detection Transformer,RT-DETR)[25],通过引入更加高效的混合编码器、IoU感知查询机制等技术,保证了检测的实时性。为了提高小目标检测的精度,Muzammul等人[26]引入切片辅助超推理(slicing aided hyper inference,SAHI)[27],有效改善了RT-DETR无人机航拍图像中的目标检测和识别能力。张储等人[28]使用Inner-IoU和OrthoBasicBlock模块优化RT-DETR模型。李亦涵等人[29]引入可变核卷积(alterable kernel convolution,AKConv)[30]来适应遥感图像中小目标的大小和形状变化。胡佳乐等人[31]再提出DySSFF模块,替换RT-DETR原有的特征融合模块,避免小目标特征信息丢失。

尽管上述方法取得了不错的进展,但仍存在一些局限性。首先,这些方法没有特别关注小目标的边界特征,未考虑到样本不平衡的问题;其次,模型的特征提取能力以及定位能力还存在可提升空间。因此,针对上述不足,本文以RT-DETR为基础提出MBFE-DETR,主要改进工作如下:

第一,设计一种增强特征提取的轻量化主干网络,该主干网络基于C2f和单头自注意力模块(CSP single-head block,CSHB),有效降低了模型的参数量,并提高网络的特征提取能力,使得密集小目标之间特征更具备区分度,降低目标被模型判定为背景的概率。

第二,提出多尺度边界特征增强协同网络(multi-scale boundary feature enhancement collaborative network,MBFECN),通过多尺度边界特征增强机制和高效特征融合策略,使模型可以感知到更细粒度的边界特征。这些边界特征能够提供更精确的目标轮廓细节,帮助模型在复杂背景中更好地检测。

第三,引入Focaler-MPDIoU以提升定位准确性,结合Focaler-IoU和MPDIoU的思想,MPDIoU不仅关注IoU的重叠程度,还额外关注框的位置匹配。Focaler-IoU通过线性区间映射重构IoU损失,从而在不同的检测任务中自适应地聚焦于不同难易程度的回归样本。

第四,构造新的分类损失函数ESVLoss,对分类损失值进行分段加权,同时对权重进行动态平滑更新,改进后的损失函数能有效解决样本不平衡的问题,对密集小目标场景有较强的适应能力。

在Visdrone2019-DET和DOTAv1.0两个数据集上进行实验,将本文算法与其他主流的检测算法进行比较,结果表明所提算法不仅有效提升了检测精度,而且在模型参数量方面也有优势。

1 MBFE-DETR算法

1.1 RT-DETR算法简介

RT-DETR是首个实时端到端目标检测模型,能够

在保证检测精度的同时显著提升推理速度,并克服传统YOLO和DETR检测器的局限性,尤其是非极大值抑所引发的延迟问题。RT-DETR通过引入高效混合编码和不确定性最小化查询选择,在实时目标检测场景中取得了优异表现。高效混合编码器通过解耦多尺度特征的内部交互和跨尺度融合,显著减少计算冗余的同时保留了特征的完整性。不确定性最小化查询选择则通过优化特征的不确定性,选取高质量的初始查询用于解码器,从而在分类与定位两方面均取得更佳的性能。RT-DETR结构如图1所示。

1.2　MBFE-DETR算法结构

针对无人机视角下由于背景复杂、小目标比例较高且样本不平衡等问题,本文选用RT-DETR作为基线模型进行改进。MBFE-DETR主要从四个方面对RT-DETR模型进行了改进。首先,通过C2f和CSHB模块构建了更高效的主干网络。其次,构建了以多尺度边界特征增强机制为核心的MBFECN,使模型可以感知到更细粒度的边界特征。采用Focaler-MPDIoU,能够在检测任务中自适应地聚焦于不同难易程度的回归样本,以提升模型在复杂场景下的定位准确性。最后,构造新的分类损失函数ESVLoss,以解决样本不平衡的问题,改进后的模型结构如图2所示。

1.3　主干改进

传统ResNet18主干网络作为经典的特征提取架构,虽然结构简洁且在多种计算机视觉任务中应用广泛,但面对无人机航拍场景下的小目标检测任务时存在一定的局限性。首先其连续的下采样操作会导致空间分辨率降低,使得小目标的特征表示在深层网络中逐渐减弱甚至丢失。并且BasicBlock结构主要依赖有限感受野的卷积操作,难以捕获目标与周围环境的长距离依赖关系,缺乏对全局上下文的有效建模。最后,ResNet18具有相对冗余的参数量,不利于模型在无人机设备上的部署。

针对上述问题,本文提出了基于C2f和CSHB模块的轻量化主干网络。在浅层特征提取中,C2f通过特征转换、分支处理和特征融合等操作,有效保留了空间细节信息,防止小目标特征在深层网络中丢失。在深层特征处理中,CSHB模块巧妙融合了局部卷积和全局自注意力机制,弥补了ResNet18中缺乏全局上下文建模的不足,使网络能够同时捕获局部细节和全局关系。同时,CSHB模块中的分支结构创建了多条梯度流通路径,缓解了深层网络训练中的梯度衰减,增强了模型的特征学习能力,进而提升了对小目标的检测能力。

在CSHB模块中,首先通过1×1卷积调整特征图的

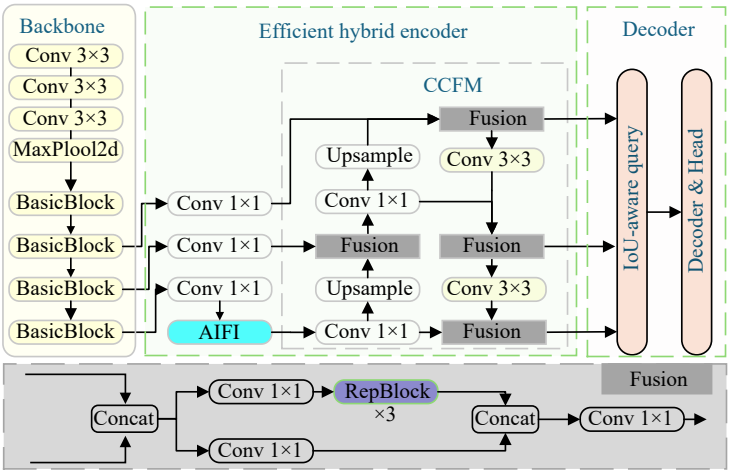


图1　RT-DETR网络结构图

Fig.1　Network structure diagram of RT-DETR

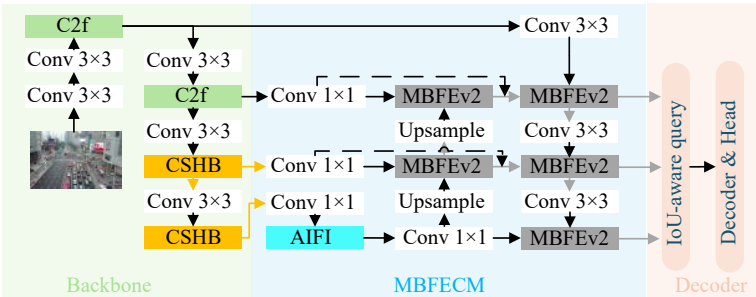


图2　MBFE-DETR网络结构图

Fig.2　Network structure diagram of MBFE-DETR

通道数,增强特征的表达能力。然后使用Split操作将特征图划分成两个部分,残差分支直接保留输入特征,确保输入的重要信息不会在主分支中丢失。主分支通过 n 个串联的SHSABlock[32]进行特征提取,并保持残差连接,在减少参数量和计算量的同时获得更丰富的梯度流。随后,将多个并行的梯度流分支通过Concat进行拼接,从而整合更丰富的梯度信息,进一步强化特征表达能力。最后,通过1×1卷积对拼接后的特征图进行处理,得到CSHB的最终输出,其结构如图3所示。

SHSABlock的核心作用是通过局部卷积和全局注意力相结合的方式,捕捉局部和全局特征。首先通过深度卷积捕获像素周围的局部纹理和结构信息。其次,通过基于部分通道卷积的单头自注意力捕获长距离依赖关系和全局上下文。最后,通过前馈网络对提取的特征进一步扩展和变换,提升特征的非线性表达能力。残差连接在每个模块中保留原始信息,防止梯度消失,并提升训练的稳定性。SHSABlock在保留Transformer的强大建模能力的同时,减少了多头注意力带来的计算开销。其中,基于部分通道卷积的单头自注意力仅在输入通道的一部分($Cp=rC$)上应用单头注意力层进行空间特征聚合,而保持剩余通道不变, r 默认设置为1/4。单头自注意力表示如下所示:

$$SHSA(X)=Concat(\tilde{X}_{att},X_{res})W^{O} \tag{1}$$

$$\tilde{X}_{att}=Attention(X_{att}W^{Q},X_{att}W^{K},X_{att}W^{V}) \tag{2}$$

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Softmax(QK^{T}/\sqrt{d_{qk}})V \tag{3}$$

$$X_{att},X_{res}=Split(X,[Cp,C-Cp]) \tag{4}$$

其中, W^{Q} 、W^{K} 、W^{V} 和 W^{O} 是投影权重, dqk 是查询和键的维(默认为16)。选取特征图前 r 个通道作为整个特征图的代表以保持内存访问的一致性。此外,SHSA的最终投影应用于所有通道,而不仅仅是前 Cp 个通道,确保注意力特征能够有效传播到剩余通道。

总的来说,CSHB是一个集局部感知和全局建模于一体的模块,充分发挥了注意力机制和卷积的优势。单头自注意力机制通过计算特征图内所有空间位置之间的关系,增强了网络对目标边界的感知能力,使得密集小目标边界特征得到强化,进而减少了相邻目标特征的混淆。残差分支结构则保留了原始特征信息,确保在特征转换过程中不丢失小目标的独特特征,同时多条梯度流通路径增强了特征表示的多样性,使得不同目标即使在视觉特征相似的情况下也能维持区分性。实验表明,本文提出的基于C2f和CSHB的主干网络不仅提高了检测精度,还降低了模型复杂度,在特征提取效率和检测精度方面均优于ResNet18。

1.4　多尺度边界特征增强协同网络

在无人机图像的目标检测任务中,小目标通常分布密集且存在相互遮挡现象,从而导致这些目标的边缘比较模糊,如何有效保留这些目标的边界特征是检测任务的主要挑战之一。因此,本文提出多尺度边界特征增强协同网络MBFECN,如图4(a)所示,该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。

其中,边界增强模块BE的主要目的是通过提取高频边界特征,增强特征图中的细节信息,如图4(d)所示。BE模块首先使用平均池化对输入特征图进行处理,提取其低频特征以平滑原始特征图。随后将原始输入特征图与平滑后的特征图相减,得到增强的边界特征,这种操作可突出特征图中的边缘和细节信息。然后对提取的边界特征使用卷积层和Sigmoid激活函数进行进一步处理,将边缘信息的范围压缩至[0,1]。最后,将处理后的边界特征与原始输入特征图相加,从而保留原始信息并增强边界特征,生成增强后的输出。给定 x 为输入特征图, $AvgPool$ 为平均池化, σ 为激活函数, $Conv$ 表示卷积操作,则BE模块的数学表达如式(5)所示:

$$BE(x)=x+\sigma(Conv(x-AvgPool(x))) \tag{5}$$

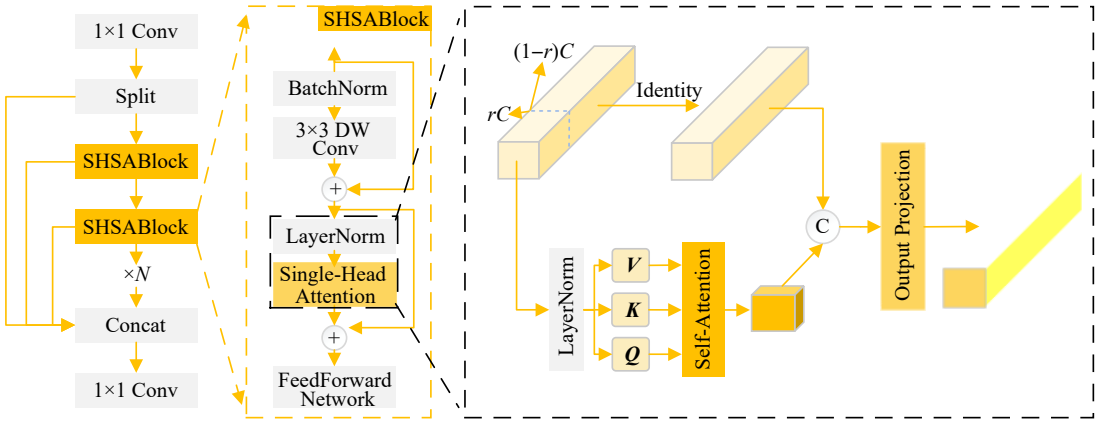


图3　CSHB模块结构图

Fig.3　Structure diagram of CSHB module

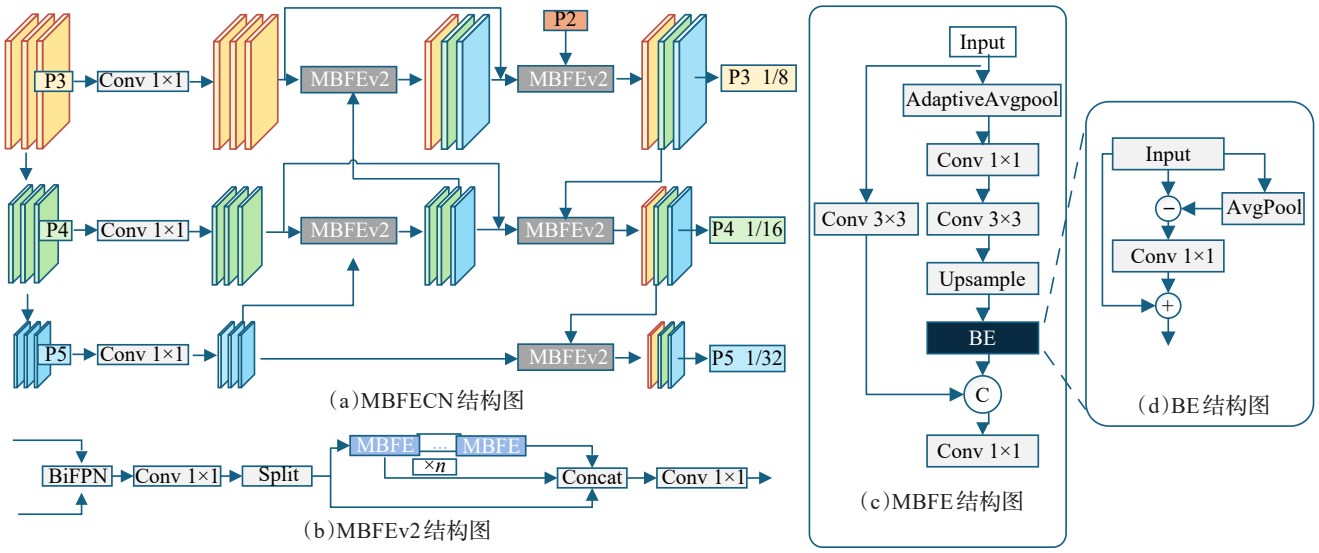


图4　MBFECN结构图

Fig.4　Structure diagram of MBFECN

在BE模块的基础上，MBFE主要通过对特征图进行多尺度处理和边界特征增强，进一步提高了对不同尺度目标的检测能力，如图4(c)所示。MBFE模块首先使用4个自适应平均池化将输入特征图调整到指定输出尺寸(3×3、6×6、9×9、12×12)，以生成不同尺度的下采样特征图。每个尺度特征图会通过一个1×1卷积实现通道压缩，并通过一个3×3卷积进一步提取局部特征。然后将每个多尺度特征图恢复至原始分辨率，并通过BE模块对其进行边界特征增强。同时，对原始输入特征图进行一次3×3卷积，以提取局部特征作为全局参考。最后，将所有增强后的多尺度特征与局部特征拼接，通过卷积层进一步融合，生成最终的增强特征。MBFE模块的数学表达如下所示：

$$F_{\mathrm{ms}}^{i}=BE(Upsample(Conv_{3\times3}(Conv_{1\times1}(AdaptiveAvgPool(x)))))\tag{6}$$

$$F_{\mathrm{local}}=Conv_{3\times3}(x)\tag{7}$$

$$F_{\mathrm{out}}=Conv([F_{\mathrm{local}},F_{\mathrm{ms}}^{1},F_{\mathrm{ms}}^{2},F_{\mathrm{ms}}^{3},F_{\mathrm{ms}}^{4}])\tag{8}$$

在MBFE的基础上，为了进一步减少计算量，本文结合CSPNet[33]中的跨阶段分离思想，提出了改进版本MBFEv2，如图4(b)所示。该模块首先将来自不同层的同一尺度特征图融合，随后将特征图分成两部分，一部分保留全局信息，另一部分作为输入送入MBFE模块以提取增强的多尺度边界特征。随后，将分割后的两部分特征拼接在一起，以便同时利用全局信息和边界增强特征。最后，通过卷积对拼接后的特征进行融合，从而生成更强的特征表达。

最终，输入到检测头的P3和P4特征图采用了跳跃连接，以结合主干网络中得到的特征图和特征融合网络中得到的对应尺度特征图，然后这些特征通过MBFEv2在通道维度上进行融合，进一步增强了特征表达能力，使无人机航拍图像的检测更为准确。MBFECN有效解决了无人机航拍图像中小目标的挑战，在保持边界细节和提高检测精度方面取得了显著效果。

1.5　Focaler-MPDIoU

在目标检测任务中，基于IoU的损失函数可有效衡量检测模型的定位性能。然而，传统的边界框回归方法(如GIoU[34]、CIoU[35])往往忽略了边界框位置匹配以及样本难易程度对回归结果的影响。因此，本文采用Focaler-MPDIoU定位损失函数，结合了Focaler-IoU[36]和MPDIoU[37]两者的思想，以提升边界回归精度并增强目标检测性能。

其中，Focaler-IoU通过对原有IoU损失进行线性区间映射重构，使其能够在不同的检测任务中，根据目标的难易程度自适应地聚焦于不同类型的回归样本。MPDIoU则在关注IoU重叠程度的基础上，增加了对边界框位置匹配的约束，即通过计算预测框与真实框两对角点的距离来度量位置偏差，从而在复杂场景中取得更优表现。如式(9)~(11)所示：

$$MPDIoU=\frac{A\cap B}{A\cup B}-\frac{d_1^2}{w^2+h^2}-\frac{d_2^2}{w^2+h^2}\tag{9}$$

$$L_{\mathrm{MPDIoU}}=1-MPDIoU\tag{10}$$

$$L_{\mathrm{MPDIoU}}=1+\frac{d_1^2+d_2^2}{h^2+w^2}-IoU\tag{11}$$

其中，A为真实框，B为预测框，d_1^2表示真实框与预测框左上角的欧几里得距离平方，d_2^2表示右下角的距离平方，w^2+h^2是输入图像对角线的平方，用于归一化距离值，IoU表示预测框与真实框的交并比，定义如图5所示。

MPDIoU的损失函数由三部分构成：(1)常数项1用于确保损失值为正；(2)距离项衡量预测框与真实框

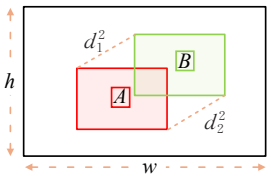


图5　MPDIoU损失函数定义图

Fig.5　Definition diagram of MPDIoU loss function

之间的位置偏差,距离越小说明匹配越好;(3)IoU项衡量预测框与真实框的重叠程度,重叠越大损失越小。通过对几何距离与IoU的综合优化,MPDIoU使预测框在关注重叠率的同时,对位置和形状能实现更精细的调整。

在此基础上,将Focaler-IoU与MPDIoU相结合,得到了Focaler-MPDIoU损失函数。Focaler-MPDIoU不仅充分考虑样本难度的差异,而且融入了边界框位置匹配因素,因而对于小目标、密集目标以及复杂背景具有更好的适应性。其最终损失函数定义如式(12)~(14)所示:

$$IoU^{\text{focaler}}=\begin{cases}0, IoU<d\\ \dfrac{IoU-d}{u-d}, d\leqslant IoU\leqslant u\\ 1, IoU>u\end{cases} \tag{12}$$

$$L_{\text{Focaler-IoU}}=1-IoU^{\text{focaler}} \tag{13}$$

$$L_{\text{Focaler-MPDIoU}}=L_{\text{MPDIoU}}+IoU-IoU^{\text{focaler}} \tag{14}$$

通过重构原有IoU损失,Focaler-MPDIoU能够灵活地对不同回归样本加以区分性地关注,即根据两个可动态调节的缩放因子 d 和 u ,对IoU较低的困难样本施加更大的学习压力,而对IoU较高的简单样本则减轻压力,这种梯度重分配策略能够使模型更加关注那些难以学习的样本,如小尺寸目标、部分遮挡目标或形状复杂目标。相比于传统IoU损失,Focaler-MPDIoU还通过引入对角点距离约束,增强了对目标位置的建模能力,尤其是在处理小目标时,由于小目标的边界框尺寸较小,传统IoU损失对位置微小偏移不够敏感,而Focaler-MPDIoU中的距离项对这种偏移提供了更直接的惩罚,使得模型能够学习到更精确的位置参数。这对无人机航拍图像中常见的小目标检测尤为重要,因为在这种场景下,目标尺寸小、数量多、分布密集,对定位精度要求高。

1.6　ESVLoss

在RT-DETR中所采用的VarifocalLoss,其权重通过预测得分与真实IoU共同确定,并在训练过程中保持不变。然而,静态权重在面对IoU分布的动态变化及样本类别不平衡(如无人机目标检测中的小目标与困难样本)时,往往表现不足。为了解决VarifocalLoss在动态适应IoU分布变化、优化困难样本以及关注小目标上的局限,本文设计一种新的分类损失函数ESVLoss。ESVLoss借助SlideLoss[38]动态加权的思想,对损失值进行分段加权调整,并根据 τ_{auto} 和样本真实分数的分布特点灵活分配权重。同时,结合指数移动平均机制EMA对权重进行动态平滑更新,使模型更具适应性,其核心损失函数如式(15)~(16)所示:

$$BCE(p_i,q_i)=-[q_i\ln(\sigma(p_i))+(1-q_i)\ln(1-\sigma(p_i))] \tag{15}$$

$$\mathcal{L}_{\text{EMASV}}=\frac{1}{N}\sum_{i=1}^{N}w_i\cdot BCE(p_i,q_i) \tag{16}$$

其中, $BCE(p_i,q_i)$ 为预测分数 p_i 和真实分数 q_i (IoU)之间的二元交叉熵, N 代表的是批次中的样本数, $\sigma(p_i)$ 是预测的Sigmoid值,调制权重 w_i 会通过SlideLoss的动态机制对损失值进行分段动态加权,即根据真实分数 p_i 与自动更新的 τ_{auto} 之间的关系自适应分段加权计算损失。具体权重计算分为三个区域,如式(17)所示:

$$w_i=\begin{cases}1, q_i\leqslant\tau_{\text{auto}}-0.1\\ \mathrm{e}^{1-\tau_{\text{auto}}}, \tau_{\text{auto}}-0.1<q_i<\tau_{\text{auto}}\\ \mathrm{e}^{1-q_i}, q_i\geqslant\tau_{\text{auto}}\end{cases} \tag{17}$$

在该过程中, τ_{auto} 表示所有样本的平均IoU,加权函数将 q_i 小于 $\tau_{\text{auto}}-0.1$ 的样本的权重置为1,将 q_i 在 $\tau_{\text{auto}}-0.1$ 和 τ_{auto} 的范围内的权重置为较高的 $\mathrm{e}^{1-\tau_{\text{auto}}}$,而将 q_i 大于 τ_{auto} 的样本置为较小的 e^{1-q_i} ,会随着 q_i 的增加逐渐减小。ESVLoss在不同区间对应的权重分布如图6所示。

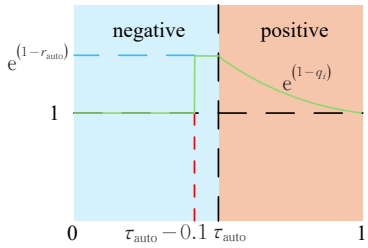


图6　ESVLoss权重分布图

Fig.6　Weight distribution diagram of ESVLoss

此外,为进一步提升模型对复杂场景的适应能力,ESVLoss中加入了EMA机制,对SlideLoss中的 τ_{auto} 值进行动态更新,如式(18)所示:

$$\tau_{\text{auto}}^{(t)}=d^{(t)}\cdot\tau_{\text{auto}}^{(t-1)}+(1-d^{(t)})\cdot IoU_{\text{current}} \tag{18}$$

其中, $\tau_{\text{auto}}^{(t)}$ EMA动态更新后的阈值, IoU_{current} 是当前批次计算出的 τ_{auto} 值, $d^{(t)}$ 是EMA的衰减因子,表示对历史值和当前值的平衡,取值范围为[0,1],会随训练过程动态变化,从而更好地融合长期与短期的信息,如式(19)所示:

$$d^{(t)}=d_{\text{base}}\cdot\left(1-\mathrm{e}^{-\frac{t}{\tau}}\right) \tag{19}$$

其中, d_{base} 是基础衰减率, t 表示更新步骤, τ 的选择会影响EMA机制对新旧数据的敏感度。通过这种设计,随着训练步数增加, $d^{(t)}$ 的值会不断逼近 d_{base} ,在平滑

更新的同时保证其渐进稳定性。

ESVLoss在动态加权的基础上引入平滑机制，可有效应对目标检测中的样本不平衡问题，尤其在小目标检测方面展现了更强的适应性。

2　实验结果

2.1　数据集

为评估本文提出算法在无人机航拍图像检测领域的提升，选用VisDrone2019-DET[39]和DOTAv1.0[40]两个数据集进行实验。

VisDrone2019-DET数据集由天津大学机器学习与数据挖掘实验室通过无人机收集，包含不同环境背景和气象条件下的场景，目标尺寸变化范围大，对算法的鲁棒性需要更高要求。数据集共含有8 629张图片，本文采用其中6 471张图片作为训练集，548张图片作为验证集，1 610张图片作为测试集。该数据集标注了行人、自行车、汽车、卡车、三轮车等类别。

DOTAv1.0数据集是一个广泛用于目标检测的航拍图像数据集，共计2 806幅航拍图片，标注了飞机、大型车辆、小型车、直升机等目标。

2.2　实验环境及参数配置

本文实验均在以下环境中进行：Ubuntu22.04.4操作系统，配备NVIDIA A40 GPU；软件环境：CUDA 12.1、Python 3.9.12和PyTorch 2.0.1。实验中所有模型均不采用预训练权重，具体实验环境参数如表1所示。

表1　实验参数设置

Table 1　Experimental parameter settings

编号	参数	设置
1	epochs	300
2	patience	50
3	batch	4
4	imgsz	640×640
5	*lr* 0	0.000 1
6	*lrf*	1
7	momentum	0.9
8	weight decay	0.000 1
9	optimizer	AdamW

2.3　实验评价指标

为了全面验证本文算法的性能，本研究选择准确率P（precision）、召回率R（recall）、平均精度AP（average precision）、平均精度均值mAP（mean average precision）、参数量（parameters）和GFLOPs作为实验评估指标。计算公式如下：

$$P=\frac{TP}{TP+FP},R=\frac{TP}{TP+FN} \tag{20}$$

$$AP=\int_0^1 P(R)\mathrm{d}R,mAP=\frac{1}{m}\sum_{i=1}^{m}AP_i \tag{21}$$

其中，TP表示被模型正确预测为正类的样本数量；FP表示被模型错误预测为正类的样本数量；FN表示被模型错误预测为负类的样本数量；AP用于衡量单个类别的检测性能。mAP是所有类别的平均精度（AP）的均值，用于衡量多类别目标检测任务的整体性能。

2.4　对比实验

2.4.1　VisDrone2019数据集上的对比实验

VisDrone2019-DET数据集背景复杂，遮挡严重并包含大量的小目标，检测难度较大。将本文算法与YOLO系列算法，SSD、RetinaNet[41]、Faster-RCNN等算法，Swin Transformer[42]等基于Transformer的算法，以及改进算法YOLOv7+Bytetrack、STD-DETR[43]、MSM-DETR[44]、ADE-YOLO[45]进行了对比实验，结果如表2所示。

相较于表2中其他性能优异的多种主流算法，本文提出的MBFR-DETR算法在VisDrone2019数据集上取得了较高的性能优势，P、R指标分别为64.2%、49.7%，关键性能指标mAP_{50}和$mAP_{50:95}$分别达到51.6%、32.1%，明显超越了基线算法RT-DETR-R18和其他对比算法。在参数量、计算量相当的情况下，与YOLO11m相比，MBFE-DETR的精度指标mAP_{50}高于YOLO11m算法7.2个百分点。相较于SSD、RetinaNet、Faster-RCNN检测算法，MBFE-DETR的mAP_{50}和$mAP_{50:95}$效果更为显著，并且模型的计算成本相对更低。与Deformable DETR相比，mAP_{50}提升8.5个百分点，$mAP_{50:95}$提升5.0个百分点，模型参数量和计算量分别降低了46.3%、65.8%。

在保持优秀检测性能的同时，MBFE-DETR还具备较轻量的模型结构，参数量和计算量分别为15.6×10^6和67.2 GFLOPs，尽管与RT-DETR-R18相比计算量略增，但与RT-DETR-R50相比，mAP_{50}提升1.8个百分点，参数量下降62.8%，计算量下降48.2%，大幅减少了计算资源的消耗。结果表明，本文提出的MBFE-DETR算法具有较高的检测精度和较轻量的模型结构，在无人机航拍目标检测任务中具有巨大的实际应用潜力。

2.4.2　可视化分析

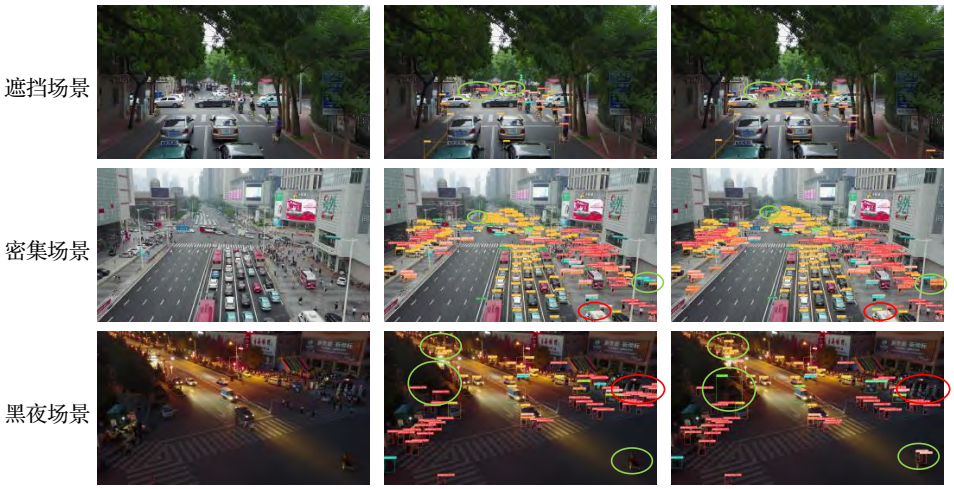
为直观验证本文算法在无人机航拍目标检测任务中的性能优势，对VisDrone2019-DET数据集中的典型场景进行检测结果可视化和热力图可视化，从多个角度展示模型的有效性。图7展示了本文提出的MBFE-DETR与基线算法在遮挡场景、密集场景以及黑夜场景下的检测对比结果，其中绿色圆圈为漏检对比区域，红色圆圈为误检对比区域。

从遮挡场景中可以观察到，本文算法成功检测到两侧道路上被遮挡的行人和面包车。在密集场景中，本文算法对道路上密集分布的小目标车辆展现出更好检测能力，同时避免了RT-DETR在下侧红色圆圈区域将运

表2　对比实验

Table 2　Comparison experiment

网络模型	P/%	R/%	mAP_{50}/%	$mAP_{50:95}$/%	Parameters/10^6	GFLOPs
SSD	21.1	35.8	24.0	11.9	12.3	63.2
RetinaNet	23.5	37.9	26.5	12.4	19.8	93.7
QueryDet	41.1	33.4	31.6	17.4	18.9	44.3
Faster-RCNN	45.3	33.8	33.2	17.0	41.2	206.7
Swin Transformer	—	—	35.6	20.6	34.2	44.5
YOLOv5m	50.3	37.9	36.3	19.2	21.2	48.3
YOLOv5l	45.1	35.2	38.7	24.3	45.9	108.4
YOLOv8m	55.7	44.3	40.9	24.3	25.8	78.7
YOLOv7	54.1	43.6	42.8	22.5	37.2	103.3
Deformable DETR	—	—	43.1	27.1	29.0	196.0
YOLOv10m	53.8	42.6	44.0	26.6	15.3	58.9
YOLO11m	54.1	43.1	44.4	27.3	20.0	67.7
YOLO11l	55.2	43.7	45.0	28.0	25.2	86.6
YOLOv8l	57.4	45.3	45.7	28.1	43.6	165.2
YOLOv7+Bytetrack	57.1	46.5	45.8	26.3	34.2	—
YOLOv10l	57.6	44.6	46.3	28.4	24.3	120.0
ADE-YOLO	—	44.7	47.6	29.5	**7.8**	—
RT-DETR-R18	60.8	46.4	47.7	29.2	19.9	57.0
MSM-DETR	—	—	49.5	30.6	22.2	72.9
RT-DETR-R50	63.6	48.8	49.8	30.8	41.9	129.6
STD-DETR	64.1	48.7	50.0	30.4	12.3	**37.8**
MBFE-DETR(Ours)	**64.2**	**49.7**	**51.6**	**32.1**	15.6	67.2



(a)原图　(b)RT-DETR-R18检测图　(c)MBFE-DETR检测图

图7　MBFE-DETR和RT-DETR-R18的检测结果对比

Fig.7　Comparison of detection results between MBFE-DETR and RT-DETR-R18

货车误检为公交车的错误。在黑夜场景中,本文算法不仅检测出了左上方被绿色圆圈标注的远处小型车辆以及左侧被遮挡的卡车,还检测出了右下方被绿色圆圈标注的小目标行人。

图8展示了MBFE-DETR与RT-DETR在典型人群和车辆场景下的热力图可视化对比结果。可以看出,无论在人群1中广场上密集分布的行人还是在人群2中昏暗街道背景下的行人,MBFE-DETR都成功关注到了RT-DETR漏检的多个小目标行人,证明其在密集小目标行人场景下的有效性。在车辆1场景中,MBFE-DETR能够有效识别上方绿色圆圈标注的远处模糊车辆以及右下方圆圈内的部分遮挡车辆。面对车辆2场景中密集停放的摩托车和电动车,MBFE-DETR在相同区域的热力图明显减少了漏检情况,证明其能够在密集小目标场景下区分并检测出更多的小目标。

可视化分析结果表明,本文提出的MBFE-DETR模型表现出了更强的环境适应性和检测稳定性,能有效减少漏检、误检等情况。

（a）原图　（b）RT-DETR-R18检测图　（c）MBFE-DETR检测图

图8　MBFE-DETR和RT-DETR-R18的热力图对比

Fig.8　Comparison of heat maps between MBFE-DETR and RT-DETR-R18

2.4.3　不同尺度下的AP值对比实验

为验证本文算法在不同大小尺度目标上的检测效果，进行如表3所示的实验。从表中可以看出，本文算法对小目标的检测效果最好，AP_s达到了21.2%，超过基线算法1.5个百分点，跟其他改进算法相比也有优势。

表3　不同尺度下AP值对比实验

Table 3　AP value comparison experiment at different scales　单位：%

网络模型	AP_s	AP_m	AP_l
YOLOv8m	15.4	36.0	42.5
YOLOv8l	16.0	37.8	45.8
YOLO11m	16.7	40.9	**46.5**
文献[31]	20.0	37.2	43.3
文献[46]	17.3	36.6	41.0
ADE-YOLO	19.3	38.6	42.3
RT-DETR-R18	19.7	39.7	44.1
MBFE-DETR(Ours)	**21.2**	**42.2**	43.0

2.4.4　不同定位损失函数对比实验

为验证Focaler-MPDIoU损失函数的有效性，将其与其他主流的定位损失在VisDrone2019-DET数据集上进行对比实验。定位损失函数不影响模型参数量和计算量，因此选择P、R、mAP_{50}、$mAP_{50:95}$四个指标进行比较。结果如表4所示，本文所提出的损失函数相比于其他各类损失函数在mAP_{50}和$mAP_{50:95}$指标上有更好的效果，分别达到49.1%和30.2%。证明了本文提出的Focaler-MPDIoU对无人机图像小目标检测任务场景的适配性。

表4　IoU对比实验

Table 4　Comparison experiment of IoU　单位：%

损失函数	P	R	mAP_{50}	$mAP_{50:95}$
GIoU	60.8	46.4	47.7	29.2
Focaler-GIoU	**62.7**	46.1	47.9	29.4
DIoU	60.8	47.4	48.0	29.3
CIoU	62.6	46.8	48.3	29.8
EIoU	61.4	47.2	48.8	30.0
MPDIoU	61.5	47.9	48.9	30.1
Focaler-MPDIoU	61.8	**48.1**	**49.1**	**30.2**

2.5　消融实验

2.5.1　VisDrone2019-DET数据集消融实验

为验证各个改进模块的具体效果和算法的合理有效性，在VisDrone2019-DET数据集上进行消融实验，每组实验均在相同的环境配置和参数设置下进行。使用mAP_{50}、$mAP_{50:95}$、参数量（Parameters）和计算量（GFLOPs）四个指标进行比较，进行的消融实验结果如表5所示。

从表5可以得出，设计的轻量化主干网络在降低计算量（降低12.3%）和参数量（降低32.2%）的同时，mAP_{50}提升了1.4个百分点，表明CSHB在保持特征提取能力的同时，有效降低了模型复杂度。多尺度边界特征增强协同网络使mAP_{50}提高了1.5个百分点，证明MBFECN通过结合多尺度边界特征增强机制和高效特征融合策略，有效解决原模型在保持边界细节方面的不足。相较于基线，ESVLoss在不增加参数量和计算量的情况下，使mAP_{50}和$mAP_{50:95}$分别提高0.9和0.7个百分点，这表

表5 消融实验

Table 5 Ablation experiment

Baseline	+主干改进	+MBFECN	+ESVLoss	+Focaler-MPDIoU	mAP_{50}/%	$mAP_{50:95}$/%	Parameters/10^6	GFLOPs
√					47.7	29.2	19.9	57.0
√	√				49.1	30.3	**13.5**	**50.0**
√		√			49.2	30.4	21.7	72.3
√			√		48.6	29.9	19.9	57.0
√				√	48.4	29.8	19.9	57.0
√	√	√			49.9	30.8	15.5	67.2
√	√	√	√		50.9	31.6	15.5	67.2
√	√	√	√	√	**51.6**	**32.1**	15.5	67.2

明ESVLoss通过关注小目标的特征表达,提高了无人机航拍场景中小目标的检测性能。将基线算法中的GIoU损失函数替换为Focaler-MPDIoU后,mAP_{50}提高了0.7个百分点,证明了Focaler-MPDIoU的有效性。

进一步将轻量化主干与多尺度边界特征增强协同网络相结合,mAP_{50}提升至49.9%,表明两个模块之间存在良好的协同作用。加入ESVLoss后,mAP_{50}达到50.9%。最终完整的MBFE-DETR模型,mAP_{50}和$mAP_{50:95}$相比基线算法分别提高了3.9、2.9个百分点,同时参数量降低了22.1%。综上所述,消融实验结果充分验证了各改进模块的有效性及其协同作用,表明MBFE-DETR模型实现了检测精度与模型复杂度的平衡,证明其在无人机航拍目标检测任务中具有巨大的实际应用潜力。

2.5.2 分类损失消融实验

为验证ESVLoss的有效性,进行了表6所示的实验。选择P、R、mAP_{50}、$mAP_{50:95}$四个指标进行比较。实验表明,Slide Loss+EMA的结果最差,其他组合都有一定的改进效果,而本文采用的分类损失函数效果最好,与基础损失函数VariFocal Loss相比mAP_{50}提高了0.9个百分点,有效提升了模型的检测能力,证明本文所采用损失函数的优越性。

表6 分类损失函数消融实验

Table 6 Ablation experiments on classification loss functions

单位:%

VariFocal Loss	+Slide Loss	+EMA	P	R	mAP_{50}	$mAP_{50:95}$
√			60.8	46.4	47.7	29.2
	√		61.0	46.8	47.8	29.4
	√	√	61.1	46.9	48.1	29.5
√	√		61.2	**47.0**	48.3	29.8
√	√	√	**62.0**	46.9	**48.6**	**29.9**

2.6 DOTAv1.0数据集

为了验证本文所提算法在不同数据集上的泛化性和有效性,将其与当前性能优异的多种主流目标检测算法在DOTAv1.0数据集上进行对比实验,结果如表7所示。

相比于YOLO11m,本文算法在P、R和mAP_{50}指标上分别领先3.8、7.6和6.0个百分点,同时保持了更小的参数量。与基线算法RT-DETR-R18相比,本文提出的MBFE-DETR在P、R、mAP_{50}和$mAP_{50:95}$等关键指标上分别提升了2.8、2.6、2.0和1.2个百分点,同时参数量减少了21.6%,进一步证明了MBFE-DETR算法的优越性,表明MBFE-DETR算法不仅在VisDrone2019-DET数据集上表现出色,在DOTAv1.0这一广泛使用的遥感目标数据集上同样展现了优异的性能。在保持较小模型体积和计算资源消耗的情况下,MBFE-DETR能够取得更高的检测精度。

表7 DOTAv1.0数据集对比实验

Table 7 Comparison experiment of DOTAv1.0 dataset

Model	P/%	R/%	mAP_{50}/%	$mAP_{50:95}$/%	Parameters/10^6
YOLOv10m	64.4	33.2	36.1	21.2	16.4
YOLOv8m	63.6	35.5	38.2	22.7	25.8
YOLO11m	67.6	36.9	40.3	24.0	20.0
RT-DETR-R18	68.6	41.9	44.3	26.8	19.9
MBFE-DETR	**71.4**	**44.5**	**46.3**	**28.0**	**15.6**

本文算法与原算法在DOTAv1.0数据集上的检测效果对比如图9所示,其中绿色圆圈为漏检对比区域,红色圆圈为误检对比区域。从场景1左上角和场景3中可以看出,MBFE-DETR成功识别出RT-DETR漏检的小目标车辆和小目标船舶。在场景2中,MBFE-DETR成功避免了RT-DETR将一整个主球场误检为两个的错误,验证了其在不同场景下的泛化能力和应用潜力。

3 结束语

本文主要针对无人机航拍图像目标检测任务中目标检测性能差、背景复杂、模型参数量大、样本不平衡等问题,基于RT-DETR提出一种多尺度边界特征增强下的MBFE-DETR模型,在减少模型参数量的基础上显著提高了检测准确性。首先,基于C2f结构和CSHB模块,设计了一种增强特征提取能力的轻量化主干网络,在降低模型参数量的同时提高了网络的特征表达能力。其次,通过结合多尺度边界特征增强机制和高效特征融合策略,提出了多尺度边界特征增强协同网络MBFECN,

（a）原图　（b）RT-DETR-R18检测图　（c）MBFE-DETR检测图

图9　DOTAv1.0数据集上的检测结果对比

Fig.9　Comparison of detection results on DOTAv1.0 datasets

有效解决了原模型在保持小目标边界细节方面的不足，提高了对小目标的检测能力。同时，引入Focaler-MPDIoU损失函数以提升模型在复杂场景下的定位精度，该损失函数通过考虑边界框的位置匹配关系并通过线性区间映射重构传统IoU损失，有效提高了模型的边界框回归能力。最后，设计了ESVLoss以解决样本不平衡问题，使模型对不同尺度和不同类别的目标具有更均衡的检测能力，从而提高了模型在复杂航拍场景中的整体性能。在VisDrone2019-DET和DOTAv1.0两个数据集上的实验结果表明，所提出的MBFE-DETR算法在多个评价指标上均取得了优异表现，证明在无人机小目标检测任务中具有显著优势。

在未来的研究中，将继续优化模型结构，降低其计算成本，以适应在计算资源有限和硬件限制的无人机航拍场景的应用。

参考文献：

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 580-587.

[2] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1440-1448.

[3] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.

[5] LI C Y, LI L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv:2209.02976, 2022.

[6] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021[J]. arXiv:2107.08430, 2021.

[7] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 7464-7475.

[8] WANG C Y, YEH I H, LIAO H P. YOLOv9: learning what you want to learn using programmable gradient information [J]. arXiv:2402.13616, 2024.

[9] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[J]. arXiv:2405.14458, 2024.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.

[11] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10778-10787.

[12] ZHANG S F, WEN L Y, BIAN X, et al. Single-shot refinement neural network for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4203-4212.

[13] SAHIN O, OZER S. YOLODrone: improved YOLO architecture for object detection in drone images[C]//Proceedings of the 44th International Conference on Telecommunications and Signal Processing. Piscataway: IEEE, 2021: 361-365.

[14] 李成豪, 张静, 胡莉, 等. 基于多尺度感受野融合的小目标检测算法[J]. 计算机工程与应用, 2022, 58(12): 177-182.
LI C H, ZHANG J, HU L, et al. Small object detection algorithm based on multiscale receptive field fusion[J]. Computer

Engineering and Applications, 2022, 58(12): 177-182.
[15] 聂源, 赖惠成, 高古学. 改进YOLOv7+Bytetrack的小目标检测与追踪[J]. 计算机工程与应用, 2024, 60 (12): 189-202.
NIE Y, LAI H C, GAO G X. Improved YOLOv7+Bytetrack small target detection and tracking[J]. Computer Engineering and Applications, , 2024, 60 (12): 189-202.
[16] 汤静雯, 赖惠成, 王同官. 远距离情形下的改进YOLOv8行人检测算法[J]. 计算机工程, 2025, 51(4): 303-313.
TANG J W, LAI H C, WANG T G. Improved YOLOv8 pedestrian detection algorithm for long-distance person[J]. Computer Engineering, 2025, 51(4): 303-313.
[17] TANG J W, LAI H C, GAO G X, et al. PFEL-Net: a lightweight network to enhance feature for multi-scale pedestrian detection[J]. Journal of King Saud University-Computer and Information Sciences, 2024, 36(8): 102198.
[18] WANG H, LIU J, ZHAO J, et al. Precision and speed: LSOD-YOLO for lightweight small object detection[J]. Expert Systems With Applications, 2025, 269: 126440.
[19] 李峻宇, 刘乾坤, 付莹. 融合注意力机制的红外小目标检测[J]. 航空学报, 2024, 45(14): 90-101.
LI J Y, LIU Q K, FU Y. Infrared small target detection based on attention mechanism[J]. China Industrial Economics, 2024, 45(14): 90-101.
[20] 张浩晨, 张竹林, 史瑞岩, 等. YOLO-CDC: 优化改进YOLOv8的车辆目标检测算法[J]. 计算机工程与应用, 2025, 61(13): 124-137.
ZHANG H C, ZHANG Z L, SHI R Y, et al. YOLO-CDC: improved YOLOv8 multi-scale vehicle object detection algorithm[J]. Computer Engineering and Applications, 2025, 61(13): 124-137.
[21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.
[22] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
[23] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[J]. arXiv: 2010.04159, 2020.
[24] LI F, ZHANG H, LIU S L, et al. DN-DETR: accelerate DETR training by introducing query DeNoising[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 13609-13617.
[25] ZHAO Y A, LV W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 16965-16974.
[26] MUZAMMUL M, ALGARNI A, GHADI Y Y, et al. Enhancing UAV aerial image analysis: integrating advanced SAHI techniques with real-time detection models on the VisDrone dataset[J]. IEEE Access, 2024, 12: 21621-21633.
[27] AKYON F C, ONUR ALTINUC S, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE, 2022: 966-970.
[28] 张储, 徐伟悦, 杨如雪, 等. 一种基于优化后的RT-DETR模型的红花目标检测方法和装置: 202410039910[P]. 2024-04-09.
ZHANG C, XU W Y, YANG R X, et al. A method and device for detecting red flower targets based on an optimized RT-DETR model: 202410039910[P]. 2024-04-09.
[29] 李亦涵, 张秀再, 沈涛. 一种改进RT-DETR算法的遥感图像目标检测方法及系统: 202410609716[P]. 2024-06-14.
LI Y H, ZHANG X Z, SHEN T. An improved RT-DETR algorithm for remote sensing image object detection method and system: 202410609716[P]. 2024-06-14.
[30] ZHANG X, SONG Y, SONG T, et al. AKConv: convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters[J]. arXiv:2311.11587, 2023.
[31] 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.
[32] YUN S, RO Y. SHViT: single-head vision transformer with memory efficient macro design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 5756-5767.
[33] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 1571-1580.
[34] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 658-666.
[35] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 8574-8586.
[36] ZHANG H, ZHANG S J. Focaler-IoU: more focused intersection over union loss[J]. arXiv:2401.01525, 2024.
[37] MA S, XU Y. MPDIoU: a loss for efficient and accurate

bounding box regression[J]. arXiv:2307.07662, 2023.

[38] YU Z P, HUANG H B, CHEN W J, et al. YOLO-FaceV2: a scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.

[39] ZHU P F, DU D W, WEN L Y, et al. VisDrone-VID2019: the vision meets drone object detection in video challenge results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. Piscataway: IEEE, 2019: 227-235.

[40] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018.

[41] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2999-3007.

[42] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9992-10002.

[43] 尹泽宇, 杨波, 陈金令, 等. 基于STD-DETR的轻量化小目标检测算法[J]. 激光与光电子学进展, 2025, 62(8): 146-156.
YIN Z Y, YANG B, CHEN J L, et al. Lightweight small object detection algorithm based on STD-DETR[J]. Laser & Optoelectronics Progress, 2025, 62(8): 146-156.

[44] 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 K Q, 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.

[45] 江旺玉, 王乐, 姚叶鹏, 等. 多尺度特征聚合扩散和边缘信息增强的小目标检测算法[J]. 计算机工程与应用, 2025, 61(7): 105-116.
JIANG W Y, WANG L, YAO Y P, et al. Small target detection algorithm based on multi-scale feature aggregation diffusion and edge information enhancement[J]. Computer Engineering and Applications, 2025, 61(7): 105-116.

[46] 潘玮, 韦超, 钱春雨, 等. 面向无人机视角下小目标检测的YOLOv8s改进模型[J]. 计算机工程与应用, 2024, 60(9): 142-150.
PAN W, WEI C, QIAN C Y, et al. Improved YOLOv8s model for small target detection from UAV perspective[J]. Computer Engineering and Applications, 2024, 60(9): 142-150.