

基于EIMYOLO的高分遥感图像目标检测

曹 峰¹, 曾科文¹, 李德玉¹, 罗喜召², 陶重霖³

(1. 山西大学计算机与信息技术学院, 山西太原 030006; 2. 苏州大学计算机科学与技术学院, 江苏苏州 215006;
3. 苏州科技大学电子与信息工程学院, 江苏苏州 215009)

摘 要: 高分遥感图像目标检测是遥感信息智能化处理的研究热点, 具有广泛的应用背景和重要的应用价值。相比于自然图像, 高分遥感图像目标检测面临目标朝向任意、尺度变化大、背景复杂易受干扰以及排列密集度高等诸多难点。为了进一步提升高分遥感图像目标检测算法的性能, 本文从特征融合与特征增强的角度出发, 以YOLO11为基准算法提出一种旋转框遥感图像目标检测算法EIMYOLO, 并设计了边缘特征增强、多尺度特征增强提取器和多尺度注意力机制动态融合3个即插即用的模块。边缘特征增强模块通过提取目标的边缘特征, 提高了算法对旋转目标的方向敏感度以及复杂背景下的特征提取能力。多尺度特征增强提取器和多尺度注意力机制动态融合模块, 分别从层内特征增强及层间特征融合角度出发, 提高了算法对密集目标和细长目标的检测能力。为了验证本文算法的性能, 在公共遥感数据集HRSC2016和DIOR-R上进行了实验。结果表明, 所提算法的平均检测精度分别达到了90.80%和72.40%, 优于基准算法和对比算法。

关键词: 高分遥感图像; 深度学习; 特征融合; 特征增强; 注意力机制

基金项目: 国家自然科学基金(No.62472269, No.62072291, No.62473241, No.62436004, No.62372317, No.62472300)

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2025)07-2266-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20250216

Object Detection Based on EIMYOLO for High-Resolution Remote Sensing Images

CAO Feng¹, ZENG Ke-wen¹, LI De-yu¹, LUO Xi-zhao², TAO Chong-ben³

(1. School of Information and Technology, Shanxi University, Taiyuan, Shanxi 030006, China;

2. School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China;

3. School of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215009, China)

Abstract: Object detection for high-resolution remote sensing images has become a key research area in intelligent remote sensing information processing, with extensive application scenarios and significant practical value. Unlike natural images, remote sensing images present unique challenges such as arbitrary object orientations, multi-scale variations, complex backgrounds, and densely arranged targets. To further improve the performance of high-resolution remote sensing image object detection, this paper proposes EIMYOLO, a novel rotated object detection algorithm based on YOLOv11, which incorporates innovative feature fusion and enhancement strategies. The proposed method integrates three plug-and-play modules designed to enhance feature representation and fusion. First, the Edge Feature Reinforcement Block improves orientation sensitivity and feature discriminability by extracting fine-grained edge information, especially in complex scenes. Second, the Interlayer Feature Enhancement Extractor boosts intralayer feature representation, particularly benefiting the detection of dense and elongated objects. Third, the Multi-Scale Attention Dynamic Fusion enhances inter-layer feature integration through adaptive attention mechanisms. Extensive experiments conducted on two benchmark datasets, HRSC2016 and DIOR-R, demonstrate the effectiveness of our approach, achieving state-of-the-art mean Average Precision (mAP) scores of 90.80% and 72.40%, respectively. These results confirm the superior performance of EIMYOLO over existing baseline and comparative methods.

Key words: high-resolution remote sensing images; deep learning; feature fusion; feature enhancement; attention mechanism

Foundation Item(s): National Natural Science Foundation of China (No.62472269, No.62072291, No.62473241, No.62436004, No.62372317, No.62472300)

1 引言

高分遥感图像目标检测旨在从高分遥感图像中识别和定位各种目标,是计算机视觉领域的一项关键任务,已广泛应用于环境保护、城市规划、农业管理和灾害监测等领域^[1].随着深度学习技术的兴起,基于深度神经网络的目标检测算法在高分遥感图像目标检测领域取得了显著进展,极大地提高了高分遥感图像目标检测和识别的精度和效率^[2].然而,相较于自然图像,遥感图像中的检测目标具有角度不定、类别多样、尺度多变、背景复杂以及排列密集等诸多特点,使得遥感图像目标检测任务尤为困难.而角度不定,即检测目标往往都是倾斜的,是遥感图像中检测目标的共性特点.对于倾斜目标,尤其当目标的长宽比例差别较大时,传统的水平框检测算法容易引入更多的背景噪声,从而影响目标检测的精度.旋转框目标检测算法在水平框目标检测的基础上增加了目标的方向信息,能够更精确地定位和分离目标,已成为遥感图像目标检测算法的主流.

在过去的研究中,研究者提出了多种基于深度学习的旋转框(Oriented Bounding Box, OBB)目标检测算法.依据旋转框的实现方式,这些算法可以分为两大类.第一类是基于感兴趣区域以及预设锚框的旋转框检测算法.如ROI Transformer^[3]通过Transformer架构改进感兴趣区域(Region of Interest, ROI)操作,利用自注意力机制动态捕捉全局上下文信息,构建有监督的旋转感兴趣区域学习器,从而实现高精度的旋转框目标检测.S2ANet^[4]提出特征对准模块(Feature Alignment Module, FAM),以全卷积方式实现特征对齐,并结合启发式锚框设计,将水平锚框细化为旋转锚,提升了检测中的定位精度与召回率.Oriented R-CNN^[5]通过引入旋转RoIAlign和旋转框回归头,有效建模目标方向信息,实现对任意角度目标的精确检测.第二类是基于含有旋转特征的损失函数实现的旋转框检测算法.如YOLO11-obbb基于YOLO11的骨干以及颈部网络,将检测头中的损失函数在CIoU基础上添加了ProbIoU,实现了实时旋转框目标检测.以上算法主要通过设计预设锚框或在损失函数中添加角度因素,实现对倾斜目标的检测,而未考虑倾斜目标特征的分布特性.由于倾斜目标的方向敏感性主要体现在边缘特征上,若边缘信息丢失或模糊,算法就难以正确推断目标的方向,因此如何有效地提取边缘特征是遥感图像目标检测算法面临的一个难题.

此外,遥感图像中待检测目标具有的诸多特点往往会导致检测算法难以充分提取目标的特征,这是遥

感图像目标检测算法面临的另一个难题.针对此问题,一些算法从特征融合的角度出发,将特征金字塔FPN(Feature Pyramid Network)^[6]及其变体BiFPN^[7]以及注意力机制引入目标检测算法,有效提取图像的多尺度特征.GLFE-YOLOX^[8]通过改进空间金字塔池化和跨阶段部分(Cross Stage Partial, CSP)连接层,使得特征金字塔网络更适合于提取融合特征和执行小目标检测任务.FFCA-YOLO^[9]在BiFPN的基础上加入了通道重加权拼接(Channel Reweight Concat, CRC)策略,对深层特征和浅层特征分配不同的权重,加强了多尺度融合的能力.同时,一些研究者发现注意力机制能够有效地聚焦与筛选检测目标的关键特征,因此加入注意力机制也成为一种优化特征提取的方式.CBAM(Convolutional Block Attention Module)^[10]将通道注意力机制和空间注意力机制融合,能够在不同维度上捕获特征之间的相关性,从而增强小目标和复杂场景下目标的检测能力.BiFormer^[11]利用双向路由注意力机制来构建金字塔网络架构,实现了动态稀疏注意力机制,使模型可以更好地捕捉细粒度和粗粒度的特征信息.以上算法虽然在一定程度上增强了模型的特征提取能力,但由于深层与浅层特征之间存在语义信息和空间信息不匹配的问题^[12],直接融合可能由于比重分配不合理导致有用的细节信息丢失,进而影响检测的效果^[13].

本文的动机是设计一种面向遥感图像的具有高精度和高实时性的目标检测算法.YOLO作为一种经典的单阶段目标检测算法,因其在检测速度和精度方面的良好表现,已在工业界得到广泛的应用.本文以YOLO11-obbb作为基线算法,从边缘特征提取、多尺度特征增强和动态特征融合的角度出发,提出了更适用于高分遥感图像目标检测的EIMYOLO算法.本文的主要创新点如下.

(1)提出了边缘特征增强模块(Edge Feature Reinforcement Block, EFRB),通过Sobel卷积提取检测目标的边缘特征,并采用双分支结构整合检测目标的边缘特征与空间特征,增强算法对旋转目标的方向敏感度以及复杂背景下的特征提取能力.

(2)提出了层内多尺度特征增强提取器(Interlayer Feature Enhancement Extractor, IFEE),专注于层内多尺度特征融合,通过空洞卷积获取不同大小感受野的特征,并将其与原始特征融合,使算法能够捕捉层内的细粒度特征,增强密集目标的特征表达能力.

(3)提出了层间多尺度注意力动态融合(Multi-Scale Attention Dynamic Fusion, MSADF)模块,将层间不

同尺度特征图加权融合,并经过广域注意力增强(Broad-Scope Attention Enhancement, BSAE)模块,动态调整特征的权重,优化特征融合的效果. BSAE能够实现长距离特征与局部特征的互补,提高算法对细长目标检测的准确率.

2 相关工作

2.1 遥感图像目标检测算法

当前的目标检测算法主要包括单阶段检测算法^[14-16]和两阶段检测算法^[17,18]两大类. 针对遥感图像的实时性要求,单阶段检测算法凭借较高的检测精度和检测速度备受关注. YOLO和DETR作为最具代表性的单阶段检测算法,拥有优越的检测性能和解决实时性挑战的潜力. 鉴于遥感图像的复杂性以及对实时性的要求,基于YOLO和DETR的遥感图像目标检测算法需要对原有算法进行进一步的优化和改进. Wu等人^[19]基于YOLOv8引入轻量级卷积(Squeeze-and-Excitation Convolution, SEConv)代替标准卷积,减少了网络的参数计数,还集成了针对微小物体检测的专用预测头,增强了对于小目标的检测性能. Cao等人^[20]将transformer集成到YOLOv5的backbone中来提高检测精度,并在检测头中添加CSL技术,实现了高分辨率旋转框遥感检测算法. Chen等人^[21]基于Deformable DETR (DEtection TRansformer)引入了三向自适应特征融合金字塔网络,通过残差连接集成不同输入尺度的可变形编码器,增强了遥感语义信息的提取能力. Wang等人^[22]在DETR中加入级联组注意力,并提出了用于尺度内特征交互的基于级联分组注意力的尺度内特征交互(Cascaded Group Attention-based Intra-scale Feature Interaction, CGA-IFI)模块,增强了跨尺度特征交互能力. 同时,为应对密集目标检测时出现的预测框重叠问题,研究者提出一系列旋转框遥感图像检测算法^[23-27],通过引入角度参数,改进损失函数让预测框更紧密地贴合目标轮廓,大幅减少背景区域的干扰,从而提高了目标检测的精度.

以上算法对于遥感图像的多尺度特征以及上下文特征提取欠考虑,因此在面对背景复杂以及尺度差异大的目标时,检测效果仍不理想. 本文通过相应模块设计及架构优化,有效提升了算法对复杂特征的提取能力,并通过边缘增强来提升算法的方向敏感性,进一步提高旋转目标的检测能力.

2.2 特征融合与注意力机制

特征融合通过整合遥感图像不同尺度、层次和来源的特征,强化特征表达的能力^[28]. 特征金字塔提供了多尺度特征融合的思路,该方法通过将具有更好语义信息的深层特征与更好位置信息的浅层特征结

合,增强模型对多尺度特征的提取能力. 受其启发,后续NAS-FPN^[29]、BIFPN、RepGFPN^[30]、MAFPN^[31]被相继提出,在相应的下游目标检测任务中取得了较好的检测效果,同时为特征融合提供了新的思路. Liu等人^[32]利用不同膨胀率的Atrous^[33]卷积细化多尺度信息,提高了模型捕获关键信息的能力. Liu等人^[34]提出了一种基于二阶统计量的加权策略,通过自上而下的特征聚合方式将多层次特征融合在一起,确保特征在不同场景下仍拥有较强的区分能力. 除此之外,一些研究者利用注意力机制能够聚焦重要特征以及融合上下文信息的能力,提升模型的性能,如Xie等人^[35]提出了注意力特征融合(Attention Feature Fusion, AFF)模块,指导高级特征映射学习低级特征所需要的特征,利用空间通道注意力机制学习相应权重,减少了层间特征冲突带来的影响. 喻九阳等人^[36]利用Transformer架构,设计了基于深度可分离卷积的局部空间特性强化的前馈网络,为模型提供了强大的全局感知能力. GAM^[37]和CBAM串行地融合通道注意力机制与空间注意力机制,使得局部注意力得到增强.

面对复杂的遥感图像,当前目标检测算法的特征表达能力仍有待进一步加强. 本文受上述特征融合方法启发,从层内和层间两个角度进行了特征融合. 对于单层特征图,融合不同膨胀率空洞卷积获取的层内多感受野特征,增强细粒度特征的提取. 对于多层特征图,通过融合注意力机制,为不同分辨率的特征图动态分配权重,避免将其直接融合导致语义鸿沟问题.

3 本文算法

本文选用YOLO11-obb作为基准框架,相比其他主流目标检测算法,它作为单阶段目标检测算法有着优异的检测速度,并且对于检测任务中的小目标有较优的特征提取能力. YOLO11-obb在YOLO11的基础上提出了OBB检测头,该检测头通过引入ProbIoU^[38]损失函数优化了旋转框检测任务的性能.

尽管YOLO11在目标检测领域有着较好的性能,但针对复杂的高分遥感图像,仍存在一些漏检和误检问题. 为了进一步提升高分遥感图像目标检测的精度,本文从多尺度特征提取角度出发,对YOLO11-obb算法进行了改进,提出了EIMYOLO高分遥感目标检测算法. 该算法在YOLO11-obb的Backbone中增加了一个边缘特征增强模块EFRB,它利用Sobel算子得到增强的边缘特征,并与原始提取的空间特征结合,提高了骨干网络的边界域感知能力. 为了充分利用骨干网络提取的特征,在YOLO11-obb的Neck中增加了多尺度特征增强提取器IFEE和多尺度注意力机制动态融合模块MSADF. IFEE利用空洞卷积进行层内特征增强,MSADF则将局

部注意力与全局注意力结合嵌入特征融合,提高了网络的特征表达能力. 本文算法的结构如图1所示.

3.1 边缘增强模块EFRB

基于卷积神经网络的目标检测算法是当前的主流,但卷积以及池化操作自身的特点,导致该类算法的特征提取方式更着重于图像的中心特征,而边缘区域的像素则被较少的卷积核覆盖,故造成了边缘信息的丢失. 遥感图像中待检测目标背景复杂,边缘信息能够有效提取遥感图像中目标的结构信息,从而提高检测的精度. 受此启发,本文提出了边缘特征增强模块EFRB,其结构如图2所示.

EFRB 模块将输入图像分为两条路径处理. 一条路径利用 Sobel 算子提取图像的边缘信息,通过水平及垂直方向上两个 3×3 的卷积核 $Sobel_{Conv_x}, Sobel_{Conv_y}$ 如公式(1),得到每个像素点在水平和垂直方向上的梯度幅度,根据梯度幅度变化计算出边缘强度和方向. 此外, Sobel 卷积减少了图像中的噪声,且其卷积核对称,中心卷积核设置为 0,确保了在计算梯度时不会改变图像的总亮度. 另一条路径则保留图像原始特征,以维持丰富的空间信息. 将这两个分支的输出特征进行融合,得到联合特征信息,有效结合了边缘特征与空间信息,增强了图像的细节和轮廓,提高了目标检测和类别判断的准确性.

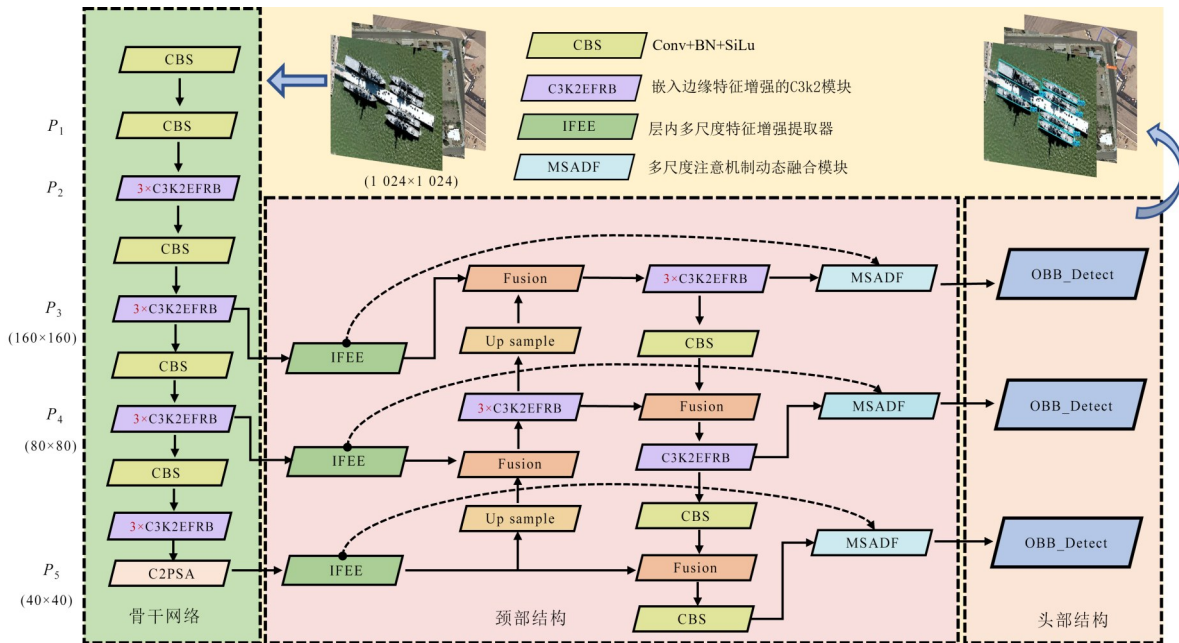


图1 EIMYOLO 总体结构图

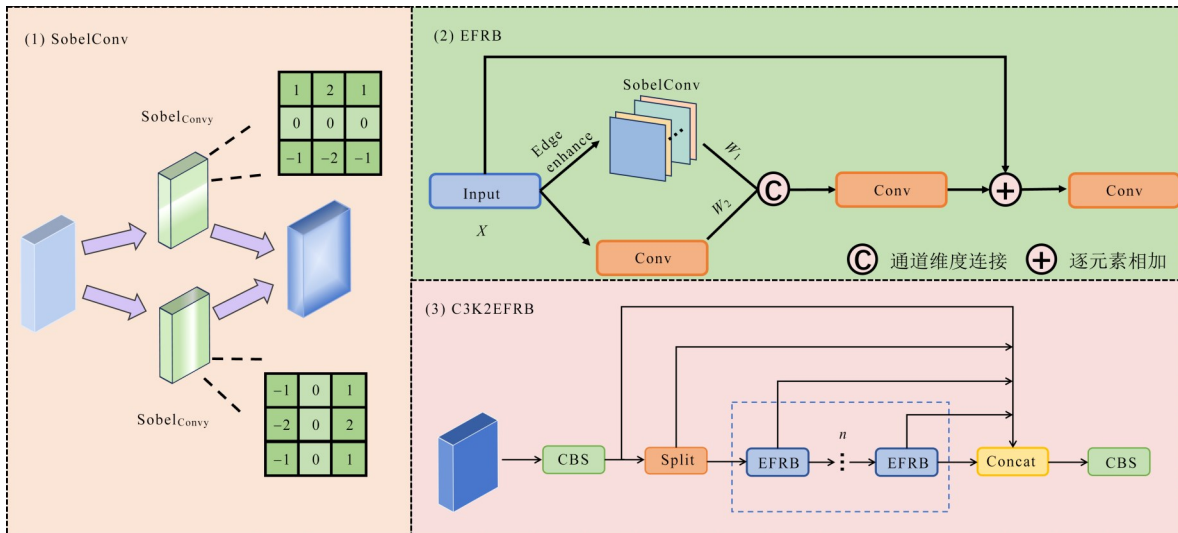


图2 SobelConv、EFRB、C3K2EFRB 结构图

$$\text{Sobel}_{\text{Conv}x} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad \text{Sobel}_{\text{Conv}y} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (1)$$

两个分支的权重分别为 W_1 和 W_2 , 如式(2)和式(3)所示, Y 为 EFRB 模块输出的特征, 如式(4)所示:

$$W_1 = f_{\text{conv}}^{3 \times 3} [\text{Sobel}_{\text{Conv}x}(X) + \text{Sobel}_{\text{Conv}y}(X)] \quad (2)$$

$$W_2 = f_{\text{conv}}^{3 \times 3} \{ \text{BN} [f_{\text{act}}(X)] \} \quad (3)$$

$$Y = f_{\text{conv}}^{1 \times 1} \{ f_{\text{conv}}^{3 \times 3} [\text{Concat}(W_1, W_2)] \} + f_{\text{conv}}^{1 \times 1}(X) \quad (4)$$

其中, $f_{\text{conv}}^{1 \times 1}$, $f_{\text{conv}}^{3 \times 3}$ 分别代表卷积核为 1×1 , 3×3 的标准卷积操作, f_{act} 是激活函数, 此处使用的是 SiLu, BN 是 Batch Normalization 层, Concat 与 + 分别是通道维度特征图连接和逐元素相加操作.

EFRB 模块旨在保留图像原始特征的同时, 通过增强边缘信息, 提升算法对图像细节的捕捉能力. 它不仅凸显了图像中的轮廓和边界, 还确保了上下文信息、目标坐标等关键元素的完整性, 使算法能够更全面、精确地理解图像内容. 本文使用 EFRB 替换 YOLO11-obb 原始 C3K2 中的 Bottleneck 部分, 生成 C3K2EFRB 模块, 利用增强的边缘特征能够更好地提高算法的方向敏感度, 对旋转目标更友好.

3.2 层内多尺度特征增强提取器 IFEE

跨层融合主要在不同分辨率的特征图上进行全局特征交互, 可能在细微特征提取上存在不足. 例如, 在一张大目标被小目标包围的高分遥感图像中, 跨层融合会因为上采样和下采样操作导致微小目标特征的丢失. 相比之下, 同一分辨率的特征融合保持了空间的一致性, 从更细粒度的角度出发, 更有效地捕捉细微特征. 因其避免了跨层融合中的上采样和下采样过程, 从而减少了信息丢失的风险. 因此, 本文提出了层内特征增强提取器 IFEE, 如图 3 所示, 通过对同一分辨率下的特征进行融合, 增强对细微特征的提取和分析能力.

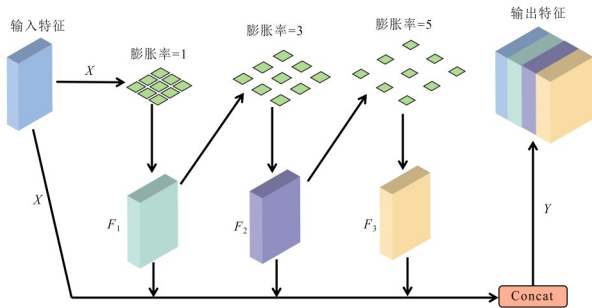


图3 IFEE模块结构图

IFEE 模块接收输入特征后, 首先将其依次通过膨胀率分别为 1、3、5 的空洞卷积层, 生成 3 个特征图: 基础感受野特征图 F_1 、宽感受野特征图 F_2 和超宽感受野

特征图 F_3 . 串行的空洞卷积操作不仅保持了特征图的原始分辨率, 还有效地捕获了不同尺度的感受野, 确保每一层特征图都能够完整地接收前一层传递的所有信息, 最大限度地减少有效特征的丢失. 随后, 这 3 个多尺度感受野特征图与原始特征图在通道维度上进行融合, 以增强特征的表达能力. 本文将 IFEE 模块集成在 Backbone 网络输出的 3 层不同尺度下采样特征图 P_3 、 P_4 、 P_5 之后 (如图 1 的颈部结构), 这一设计为传入 Neck 网络的部分提供了局部特征的增强, 并与后续的跨层多尺度特征融合操作相结合, 有效提升了算法在短距离和长距离特征提取方面的能力. IFEE 模块从输入到输出的计算过程可以用式(5)~式(8)表示:

$$F_1 = \text{AtrousConv}_1 [f_{\text{conv}}^{3 \times 3}(X)] \quad (5)$$

$$F_2 = \text{AtrousConv}_3 (F_1) \quad (6)$$

$$F_3 = \text{AtrousConv}_5 (F_2) \quad (7)$$

$$Y = f_{\text{conv}}^{1 \times 1} \{ \text{Concat} [f_{\text{conv}}^{1 \times 1}(X) + F_1 + F_2 + F_3] \} \quad (8)$$

其中, AtrousConv_1 、 AtrousConv_3 、 AtrousConv_5 分别表示膨胀率为 1、3、5 的空洞卷积, $f_{\text{conv}}^{1 \times 1}$ 、 $f_{\text{conv}}^{3 \times 3}$ 表示卷积核为 1×1 和 3×3 的标准卷积, Y 表示将原始特征图与多尺度感受野特征图沿通道维度堆叠的特征图.

3.3 多尺度注意力机制动态融合模块 MSADF

在 YOLO11 的特征融合部分以及一些改进的 FPN 结构中, 针对不同分辨率的特征图, 经过上下采样操作之后直接融合作为输出. 虽然直接融合不同尺度特征图, 能够实现语义和位置信息的初步交互, 但简单的融合方式可能造成语义鸿沟问题, 即包含更抽象语义信息的深层特征与包含更多细节信息的浅层特征具有语义差异, 导致信息融合不充分. 因此, 本文采用加权策略解决不同分辨率的特征融合问题, 提出了多尺度注意力机制动态融合模块 MSADF, 如图 4 所示. MSADF 将全局注意力机制与局部注意力机制结合, 为跨尺度特征融合动态分配权重.

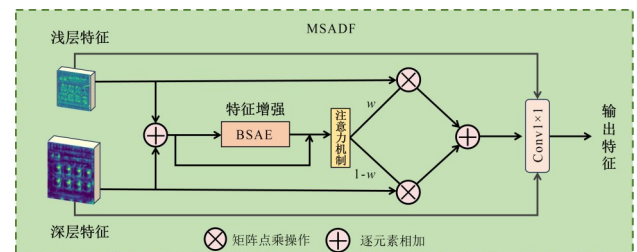


图4 MSADF模块结构图

由图 4 可知, 深层 (Deep Level) 与浅层 (Low Level) 特征图首先被送入一个名为广域注意力增强 (Broad-Scope Attention Enhancement, BSAE) 模块的子模块, 如图 5 所示. BSAE 模块包含局部注意力和全局注意力这

两个并行的注意力分支,分别负责进行局部特征 F_{Local} 和全局特征 F_{Global} 的提取. BSAE 模块可以用式(9)形式化表示:

$$F_{out} = F_{Local} + F_{Global} \quad (9)$$

全局注意力分支采用了自注意力机制,首先通过深度可分离卷积得到维度为 $W \times H \times C$ 的权重矩阵, Q 、 K 和 V 如式(10)所示:

$$Q, K, V = \text{DWConv } 3 \times 3(X) \quad (10)$$

其中, X 代表输入特征, $\text{DWConv } 3 \times 3$ 代表卷积核为 3×3 的深度可分离卷积.

然后,利用权重矩阵 Q 、 K 和 V ,计算出注意力得分,如式(11)所示,并通过 1×1 的卷积调整通道数,从而获得包含全局信息的特征图,如式(12)所示:

$$\text{Attention}(Q, K, V) = \frac{\exp\left(\frac{QK^T}{\sqrt{d_k}}\right)}{\sum_i \exp_i\left(\frac{QK^T}{\sqrt{d_k}}\right)} V \quad (11)$$

$$F_{Global} = f_{conv}^{1 \times 1}[\text{Attention}(Q, K, V)] + X \quad (12)$$

其中, $\text{Attention}()$ 代表选择的注意力机制, d_k 代表 K 的维度.

局部注意力分支首先通过通道维度的混洗操作,对输入特征图进行分组,并对每组特征执行卷积操作. 这一过程不仅增强了不同通道之间的信息交互,还进一步提升了特征的表达能力. 然后,利用全局平均池化和全局最大池化来聚合空间信息,并通过 1×1 的标准卷积进行降维和升维操作. 最后,通过 Softmax 激活函数对权重进行归一化,得到富含局部注意力的特征图,如式(13)所示:

$$F_{Local} = \frac{\exp\left\{f_{conv}^{1 \times 1}\left[\text{Avg}\left(\text{Shuffle}(X), \text{Max}\left(\text{Shuffle}(X)\right)\right)\right]\right\}}{\sum_i \exp\left\{f_{conv}^{1 \times 1}\left[\text{Avg}\left(\text{Shuffle}(X), \text{Max}\left(\text{Shuffle}(X)\right)\right)\right]\right\}} \quad (13)$$

其中, $\text{Avg}()$ 和 $\text{Max}()$ 分别代表全局平均池化和全局最大池化, $\text{Shuffle}()$ 表示通道混洗操作.

在特征动态融合阶段,将经过 BSAE 模块得到的丰富特征信息传入权重分配器,为浅层与深层特征图分配相应权重. 本文选择了通道注意力机制 CA 进行动态权重计算,通过为深层与浅层特征图分配互补的权重值,实现多尺度加权融合. MSADF 模块的主要流程可以用式(14)~式(16)表示:

$$F = X_{Low} + X_{Deep} \quad (14)$$

$$W = \text{Attention}[BSAE(F) + F] \quad (15)$$

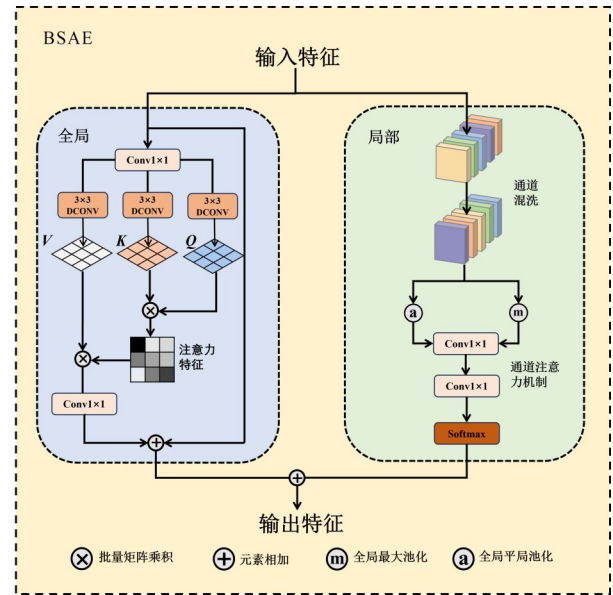


图5 BSAE 模块结构图

$$Y = f_{conv}^{1 \times 1}[W \times X_{Low} + (1 - W) \times X_{Deep} + F] \quad (16)$$

其中, X_{Low} 和 X_{Deep} 分别表示浅层与深层特征, $\text{BSAE}()$ 表示使用该模块处理, W 和 Y 分别表示权重值和输出.

本文的模型架构中, MSADF 模块集成在检测头之前, 该模块在传入检测头前, 先将来自 IFEE 的特征与深层特征进行多尺度融合, 不仅实现了层内特征的整合, 还有效融合了跨层级的特征, 形成了互补的信息流. 通过这种策略, 模型能够同时捕获局部细节和全局上下文信息, 增强特征的表达能力. 这种细致的特征融合不仅提升了模型对小目标的检测精度, 也提高了对复杂场景中目标的识别能力.

4 实验

4.1 数据集

4.1.1 HRSC2016 数据集

HRSC2016 数据集^[39]由西北工业大学在 2016 年发布, 所有图像提取自 Google Earth 中的 6 个重要港口. 该数据集总共包含 1 061 张图像以及 3 大类、27 小类共 2 976 个实例. 这些实例使用旋转的边界框进行标注, 可以进行旋转框目标检测算法的验证. 图像尺寸从 300 像素 \times 300 像素到 1 500 像素 \times 900 像素不等. 本文中训练集、验证集、测试集分别包含 436、181、444 幅图像.

4.1.2 DIOR-R 数据集

DIOR 数据集是一个专门用于光学遥感图像目标检测的大规模基准数据集, 由西北工业大学于 2022 年发布, DIOR-R^[40]在其基础上进行了旋转框标注, 它包含 23 463 幅图像和 192 472 个实例, 涵盖了 20 个不同的对象类别, 包括飞机 (AP)、机场 (AT)、棒球场 (BF)、篮

球场(BC)、桥梁(BD)、烟囱(CM)、水坝(DAM)、高速公路服务区(ESA)、高速公路收费站(ETS)、港口(HB)、高尔夫球场(GF)、地面跑道(GTF)、立交桥(OP)、船舶(SH)、体育场(STA)、储罐(STO)、网球场(TC)、火车站(TS)、车辆(VE)、风车(WM)。每幅图像被固定为800像素×800像素。本文在实验时严格遵循发布方设置,将该数据集划分为训练集、验证集、测试集3个数据集。训练集与验证集分别包含5 832和5 863张图像,一共涉及68 073个实例对象,测试集包含11 738张图像,涉及124 445个实例对象。

4.2 评价指标

本文使用 mean Average Precision (mAP) 作为主要的精度指标, mAP 的公式如下所示:

$$P = \frac{TP}{TP + FP} \quad (17)$$

$$R = \frac{TP}{TP + FN} \quad (18)$$

$$AP = \int_0^1 P(R) dR \quad (19)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (20)$$

其中, P 和 R 代表了查准率和查全率, TP (True Positive) 表示实际是正样本分类预测为正样本的数目; FP (False Positive) 表示实际是负样本分类预测为正样本的数目; FN (False Negative) 表示实际是正样本分类预测为负样本的数目; AP 表示某一类别的检测精度; N 代表类别数。同时使用指标 FPS (Frames Per Second) 表示每秒可以处理的图片数量, 用于评价算法的检测速度; 使用 FLOPS (Floating Point Operations Per Second) 表示每秒可以处理的浮点操作, 用于衡量算法在推理及训练时的计算性能; 使用 Params (Parameters) 表示模型的参数量, 用于评价模型的复杂度。

4.3 参数设置

本文实验使用的操作系统为 Ubuntu20.04, 计算平台为 CUDA12.2, 训练环境为 PyTorch2.2.1。所有实验在 NVIDIA GeForce RTX 4090 上训练, 显存为 24 GB。实验统一使用 Adam^[41] 优化器优化算法参数。初始学习率设置为 0.001, 最小学习率为 0.000 1, 训练 IoU 阈值设置为 0.7, 动量与权重衰减分别设置为 0.937 与 0.000 5, 训练 epoch 设置为 400, 早停条件设置为 50, 即 50 个 epoch 中算法没有明显提升时停止训练。

4.4 对比试验

4.4.1 HRSC2016 数据集对比分析

为了验证本文算法在背景复杂、噪声干扰和特征模糊等情况下目标检测的有效性, 在 HRSC2016 数据集上进行了对比试验, 与一些最先进的目标检测算法 Gliding Vertex^[42]、R3Det^[43]、AOPG^[44]、CPMFNET、YO-

LOv8-obb、ReDet^[45]、SARFA-Net^[46]、AFDR-Det^[47] 和 MRPENet^[48] 进行了性能对比。同时, 本文对 YOLOv9 和 YOLOv10 算法进行了改进, 使其适用于角度不定目标的检测, 也作为对比算法。改进的算法命名为 YOLOv9-obb 和 YOLOv10-obb。实验对比结果如表 1 所示。

由表 1 可知, 本文算法的 mAP 达到了 90.8%, 优于所有对比算法。与 YOLOv8-obb、YOLOv9-obb 和 YOLOv10-obb 算法相比, 本文算法的 mAP 分别提升了 4.7、2.5 和 1.5 个百分点。由此可知, 尽管 YOLO 系列算法在通用目标检测领域表现出色, 但在遥感图像目标检测任务中仍存在一定的局限性。而本文算法有效提高了遥感图像目标检测的精度, 特别是 IFEE 模块的层内特征融合与 MSADF 模块层间特征融合的互补策略, 使得所提算法能够在复杂的特征中提取到关键信息。本文算法的 FLOPS 为 30.1×10^9 , 略高于 YOLOv8-obb 和 YOLOv10-obb 算法, 远低于 RoI Trans、Oriented R-CNN、ReDet、AFDR-Det 和 CPMFNet 等算法。FPS 为 150.2, 虽略低于 YOLO 系列算法, 但明显优于对比方法中的两阶段检测方法, 可以较好地满足实时目标检测的要求。Params 相比于 YOLO 系列算法分别减少 0.2×10^6 、 2.6×10^6 、 1.2×10^6 , 而对比于其余算法, 轻量优势尤为明显。可以看出, 本文算法在拥有 YOLO 算法轻量化优势的同时, 保持着较高的检测精度。

为了进一步说明本文算法的检测效果, 选择了 HRSC2016 数据集中具有较高检测难度的 4 类检测目标, 与基线检测算法 YOLO11s 进行了对比, 部分对比结果如图 6 所示。可以看出, 对于有噪声干扰的船舶图像, 本文算法忽略了水流等噪声因素的影响, 精确定位目标本身, 精度远高于基线 59%。针对方向不定的船只图像, 本文算法优化了背景相似导致的误检, 以及角度差异大导致的漏检问题。对于背景复杂的港口图像, 基线算法受到背景更大程度的干扰, 导致检测率只有 44%, 而本文算法检测率提升了 45%。面对特征非常相似的码头与船只模糊图像, 本文算法较基线精度提升 3%, 且不存在误检。

4.4.2 DIOR-R 数据集对比分析

为了验证本文算法检测密集排列、尺度多变和类别多样目标的有效性, 在 DIOR-R 数据集上, 与 Faster RCNN、RoI Trans、DODet^[49]、PIIDet^[50] 等算法进行了对比, 对比结果如表 2 所示。

从单个类别上与其他算法对比, 本文算法在小目标类别船舶(SH)和车辆(VE)上的 mAP 分别达到了 93.4% 和 62.2%, 远远优于大多数对比算法, 证明了本文算法具有较优的小目标检测能力。此外, 本文算法在宽高比例差距较大的类别, 如桥梁(BD)、烟囱(CM)和立交桥(OP)上的精度分别为 53.5%、81.5% 和 64.2%, 优

表1 HRSC2016数据集不同算法实验结果对比表

方法	骨干网络	FLOPS/ 10^9	FPS/(帧/秒)	Params/ 10^6	mAP/%
YOLOv8-obb	CSP-DarkNet	29.4	275.7	11.4	86.10
RoI Trans	ResNet101	225	5.88	273	86.20
Gliding Vertex	ResNet101	240.2	15.0	64.9	88.20
YOLOv9-obb	CSP-DarkNet	33.4	265.3	13.8	88.30
R3Det	ResNet101	—	12.8	41.5	89.26
YOLOv10-obb	CSP-DarkNet	27.5	240.2	12.4	89.30
AOPG	ResNet101	—	12.4	—	90.34
Oriented R-CNN	ResNet101	211	21.3	41.14	90.40
SARFA-Net	ResNet50	198.1	5.8	39.22	90.42
ReDet	ReR50	80	—	—	90.46
AFDR-Det	ResNet50	235.7	—	44.34	90.51
MRPENet	ResNet50	—	13.9	—	90.60
CPMFNet	ResNet50-FPN	218	24.3	42.3	90.62
EIMYOLO(ours)	CSP-DarkNet	30.1	150.2	11.2	90.80

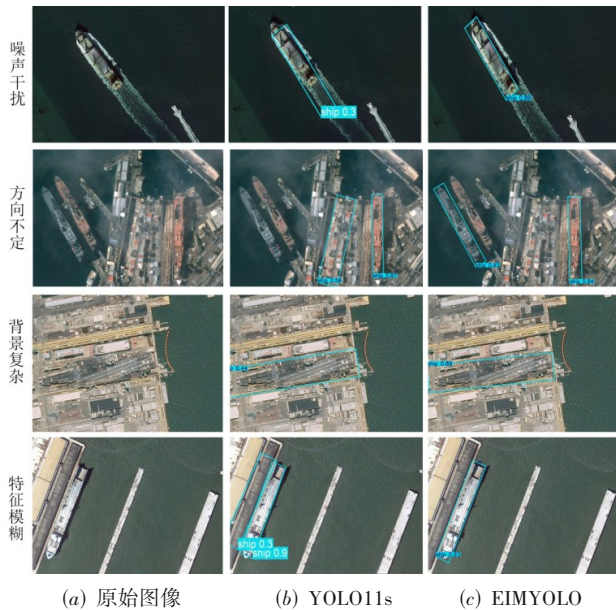


图6 HRSC2016数据集的目标检测结果对比图

于大部分对比算法。主要原因是本文算法提出的MSADF模块,具有较好地捕捉细长目标长距离特征,同时增强其局部特征的能力。针对复杂背景类别,本文算法在篮球场(BC)和港口(HB)这类包含大量重叠目标的类别上取得了90.7%和62.7%的最好结果。本文与其他版本的改进YOLO算法如YOLOv8-obb、YOLOv9-obb以及YOLOv10-obb的检测效果进行了对比。实验结果表明,本文算法在大部分类别上的检测效果表现更佳,甚至在水坝(DAM)类别上比对比算法分别高约18.9、5.0和7.4个百分点。而对于较难检测的细长目标飞机场(AT)和火车站(TS),本文算法分别优于YOLOv10-obb 4.7个百分点和8.6个百分点。由于该数据集中不同类别实例数量以及尺寸差异较大,本文算法虽

然在一些类别上的精度低于部分算法,但在多数类别上取得了最好的效果,且所有类别平均精度mAP达到了72.40%,高于所有对比算法。

DIOR-R数据集的可视化检测结果如图7所示,该图展示了20个类别的检测结果。由图7可以看出,本文算法能够较好地检测各类目标,如网球场和篮球场等类间相似度高的目标,水坝、飞机场、火车站等细长目标,以及船只和码头等密集重叠的目标。由此可见,本文算法对于复杂且类别多样的遥感图像,仍具有出色的目标检测效果。

4.4.3 与重量级基线算法的对比分析

为了进一步验证算法的性能,本文与基线算法的重量模型YOLO11m进行了对比实验,结果如表3所示。在HRSC2016数据上,本文算法在FLOPS和Params方面显著低于YOLO11m,FPS大幅提升,且mAP提高了0.8个百分点。在DIOR-R数据集上,本文算法的计算量和参数量同样低于YOLO11m,FPS提升了107.8,平均精度也提高了0.2个百分点。本文算法基于轻量模型YOLO11s进行改进,通过引入所提出的创新模块,在计算量和参数量方面进行了合理优化与调整,实验结果表明,改进后的算法在保持较低计算复杂度的同时,提升了目标检测的性能。因此,在实时目标检测任务中,表现出更强的适应性和更高的效率,能够更好地满足实际应用中对于实时性和准确性的双重需求。

4.5 消融实验

为了验证本文设计的各个模块的有效性,分别在HRSC2016和DIOR-R数据集上进行了消融实验,实验结果见表4。在HRSC2016数据集上,基准算法YOLO11s的mAP为88.1%,依次加入C3K2EFRB、IFEE和MSADF模块后,mAP分别提高了0.7、1.6和2.7个百

表2 DIOR-R数据集不同类别检测精度对比表

单位:%

方法	Faster RCNN	RoI Trans	DoDet	CPMFNet	YOLOv8	YOLOv9	PIIDet	YOLOv10	EIMYOLO(ours)
AP	62.79	63.34	63.40	71.75	91.10	85.40	79.77	88.00	89.70
AT	26.80	37.88	43.35	46.19	39.50	53.50	43.78	54.00	58.70
BF	71.72	71.78	72.11	78.81	83.30	75.90	80.04	79.90	81.30
BC	80.91	87.53	81.32	90.03	90.00	88.11	90.03	90.42	90.70
BD	34.20	40.68	43.12	45.01	47.60	42.10	47.95	50.85	53.52
CM	72.57	72.60	72.59	78.19	77.40	76.08	80.26	80.63	81.50
DAM	18.95	26.68	33.32	34.53	24.00	37.90	32.16	35.50	42.94
ESA	65.75	68.09	70.84	78.83	62.20	63.61	86.77	70.32	70.70
ETS	66.45	78.71	78.77	69.27	63.71	54.70	77.67	63.29	66.62
GF	66.63	68.97	74.15	76.58	73.62	79.71	78.46	82.54	83.10
GTF	79.24	82.74	75.47	83.29	65.10	68.41	82.10	71.30	69.67
HB	34.95	47.71	48.00	45.59	53.60	58.90	45.16	58.81	62.72
OP	48.79	55.61	59.31	58.91	61.92	60.41	58.99	62.86	64.20
SH	81.14	81.21	85.31	81.09	93.12	91.81	89.49	93.30	93.42
STA	64.34	78.23	74.04	79.81	73.20	65.50	66.14	67.92	64.52
STO	71.12	70.26	71.56	69.88	85.40	76.81	78.94	86.54	85.50
TC	81.44	81.61	81.52	88.98	92.41	89.42	89.95	90.60	90.70
TS	47.31	54.86	55.47	59.42	51.40	71.60	64.82	60.90	69.50
VE	50.46	43.27	51.86	48.58	61.80	68.32	62.88	64.42	62.21
WM	65.20	65.52	66.40	65.91	85.00	81.47	71.65	82.70	83.80
mAP	59.52	63.87	65.10	67.58	69.00	69.32	70.35	71.80	72.40

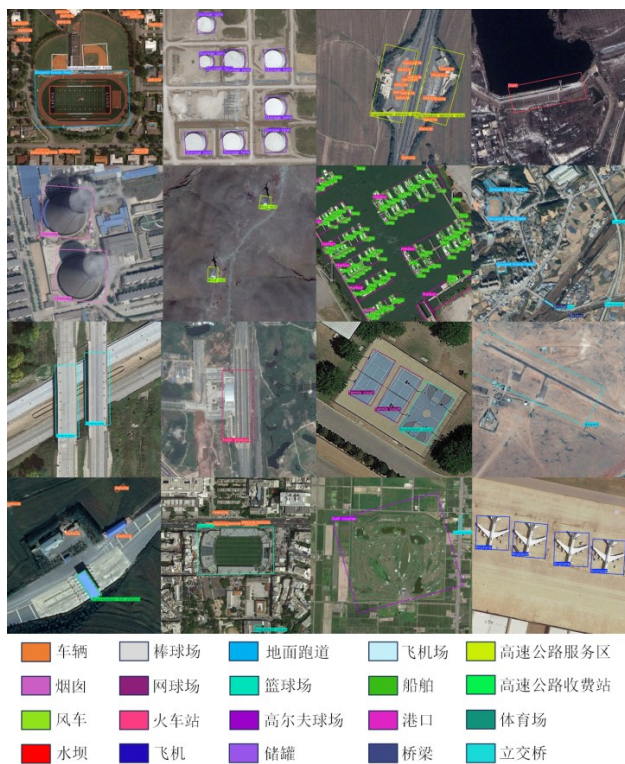


图7 DIOR-R数据集的目标检测结果图

分点. 增加各个模块后, Params 与 FLOPS 仅少量上涨, FPS 虽然低于基准算法, 但足以满足遥感图像检测的实

时性要求且准确性有一定程度的提高. 在 DIOR-R 数据集上, 依次加入 C3K2EFRB、IFEE 和 MSADF 模块后, 与基准算法 YOLO11s 相比, mAP 分别提高了 0.4、1.2 和 2.1 个百分点. 在计算量和参数量少量增加的情况下, 本文提出的模块同样有较好的精度提升效果, 同时 FPS 为 185.6, 仍满足实时目标检测要求.

4.6 注意力超参数对比实验

注意力机制是本文所提算法的一个重要的超参数. 在 MSADF 模块中, 本文基于注意力机制构建增强特征以及浅层和深层特征的多分支动态权重分配器. 为了选择最优的注意力机制, 本文进行了对比实验, 实验结果如表 5 所示. 当未引入注意力机制对各分支权重进行学习时, 模型的 mAP 为 90.00%. 当引入像素注意力机制 XA^[51] 和 EMA^[52] 时, mAP 分别提升至 90.08% 和 90.40%, 但同时带来了较大的 FLOPS 增长. 相比之下, 当引入空间注意力机制 SA^[53] 和通道注意力机制 CA^[54] 时, 在计算量与参数量仅小幅增加的情况下, 分别将 mAP 提升至 90.64% 和 90.80%. 综合精度提升与计算效率, 最终本文选择将通道注意力机制 CA 引入动态权重分配器中.

4.7 热力图可视化对比分析

为了进一步验证本文模块对于骨干网络输入图像的特征融合效果, 从 2 个数据集中选取具有识别难度的

表3 本文算法与重量模型YOLO11m的性能指标对比表

Methods	HRSC2016				DIOR-R			
	FLOPS/10 ⁹	Params/10 ⁶	FPS/(帧/秒)	mAP/%	FLOPS/10 ⁹	Params/10 ⁶	FPS/(帧/秒)	mAP/%
YOLO11m	71.3	20.88	93.7	90.0	72.0	20.92	77.8	72.2
EIMYOLO	29.9	11.20	150.2	90.8	29.0	12.40	185.6	72.4

表4 消融实验参数指标对比表

数据集	C3K2EFRB	IFEE	MSADF	Params/10 ⁶	FLOPS/10 ⁹	FPS/(帧/秒)	Precision/%	Recall/%	mAP/%
HRSC2016	×	×	×	9.8	23.0	256	85.9	84.6	88.1
	√	×	×	10.1	25.0	182.3	87.1	84.3	88.8
	√	√	×	11.1	27.9	156.7	87.3	85.2	89.7
	√	√	√	11.2	29.9	150.2	88.3	90.0	90.8
DIOR-R	×	×	×	9.9	22.4	264.7	77.7	64.2	70.3
	√	×	×	10.4	24.8	248.5	79.4	64.0	70.7
	√	√	×	12.5	28.1	207.4	82.0	65.3	71.5
	√	√	√	12.4	29.0	185.6	83.3	64.7	72.4

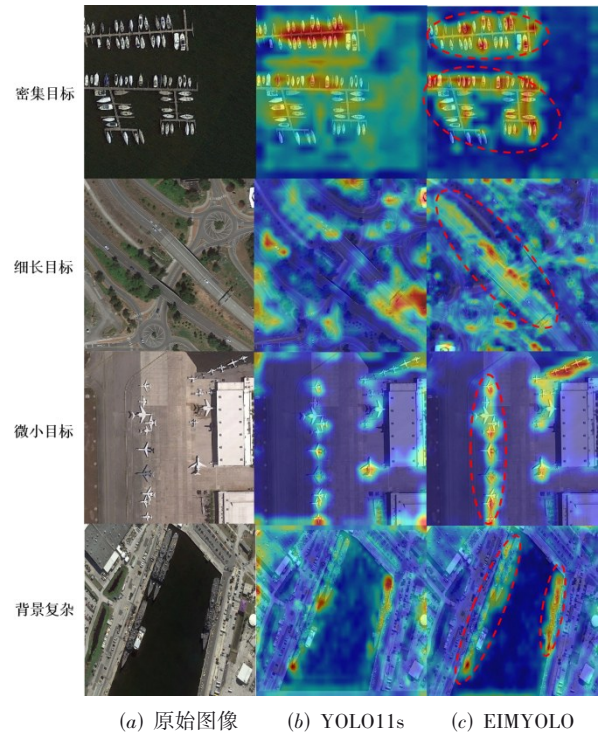
表5 注意力超参数对比表

MSADF	FLOPS/10 ⁹	Params/10 ⁶	mAP/%
Without attention	29.1	11.2	90.00
+XA	30.3	12.3	90.08
+EMA	31.8	14.2	90.40
+SA	29.4	11.3	90.64
+CA(ours)	29.9	11.2	90.80

密集、细长、微小以及背景复杂目标,进行了图像特征热力图可视化对比实验,部分结果如图8所示.可以看出,对于具有密集船只目标的码头遥感图像,基线算法YOLO11s面对密集目标时将更多注意力聚焦在码头等环境因素上,而本文算法主要将关注区域放在目标本身,对于更细微的密集船只内部也能够精准定位,避免了无关因素的干扰.面对细长立交桥遥感图像,能更好地捕捉到长距离连续特征,完整注重桥身关键特征,有效解决了基线主要关注桥头区域从而导致的目标不完整问题.面对微小飞机遥感图像,基线算法能预测目标大致位置,而本文算法在更准确地捕捉飞机目标的同时,连贯性更强,避免了目标漏检.针对复杂背景的港口遥感图像,本文算法减少了港口背景噪声干扰,注意力更好地覆盖所有待检测目标本身而不仅是船头,能够更好定位目标整体.通过以上4类热力图实验分析,体现了本文算法具有更好的目标特征捕捉能力.

5 结论

高分遥感图像的特殊性,导致传统图像目标检测算法直接应用于高分遥感图像时,其检测的性能往往难以令人满意.本文从特征融合与特征增强的角度出发,提出了一种基于YOLO11改进的高分遥感图像旋转框目标检测算法EIMYOLO.首先,本文设计了EFRB模块,并与C3K2相结合,形成了C3K2EFRB结构.该结构



(a) 原始图像 (b) YOLO11s (c) EIMYOLO

图8 检测目标特征的热力图对比

在保留丰富空间信息的同时,强化了边缘信息的提取能力,从而更好地捕捉旋转目标在方向多样性和复杂背景场景下的特征.其次,提出了IFEE模块,通过空洞卷积提取并融合骨干网络中单层特征的不同感受野,有效提升了算法对细微特征的捕捉能力,增强密集目标的特征表达能力.此外,还引入了MSADF模块,该模块将全局注意力与局部注意力融合用于动态调整特征融合时的权重,在保持全局结构感知的基础上捕捉关键的局部细节信息,实现长距离特征与局部特征的互补.本文算法为进一步提高具有挑战性的高分遥感图

像目标检测任务的检测性能提供了一种新思路。

参考文献

- [1] 刘颖, 刘红燕, 范九伦, 等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-601.
LIU Y, LIU H Y, FAN J L, et al. A survey of research and application of small object detection based on deep learning[J]. *Acta Electronica Sinica*, 2020, 48(3): 590-601. (in Chinese)
- [2] 陈立, 张帆, 郭威, 等. 面向遥感图像的多阶段特征融合目标检测方法[J]. 电子学报, 2023, 51(12): 3520-3528.
CHEN L, ZHANG F, GUO W, et al. Multi-stage feature fusion object detection method for remote sensing image[J]. *Acta Electronica Sinica*, 2023, 51(12): 3520-3528. (in Chinese)
- [3] DING J, XUE N, LONG Y, et al. Learning RoI transformer for detecting oriented objects in aerial images[EB/OL]. (2018-12-01)[2025-03-17]. <https://arxiv.org/abs/1812.00155>.
- [4] HAN J M, DING J, LI J, et al. Align deep features for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 5602511.
- [5] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 3500-3509.
- [6] 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. Piscataway: IEEE, 2017: 936-944.
- [7] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10778-10787.
- [8] GU Q, HUANG H S, HAN Z G, et al. GLFE-YOLOX: Global and local feature enhanced YOLOX for remote sensing images[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 2516112.
- [9] ZHANG Y, YE M, ZHU G Y, et al. FFCA-YOLO for small object detection in remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5611215.
- [10] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block attention module[C]//Computer Vision - ECCV 2018. Cham: Springer, 2018: 3-19.
- [11] ZHU L, WANG X J, KE Z H, et al. BiFormer: Vision transformer with bi-level routing attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 10323-10333.
- [12] 姚艳清, 程焱, 谢星星, 等. 多分辨率特征融合的光学遥感图像目标检测[J]. 遥感学报, 2021, 25(5): 1124-1137.
YAO Y Q, CHENG G, XIE X X, et al. Optical remote sensing image object detection based on multi-resolution feature fusion[J]. *National Remote Sensing Bulletin*, 2021, 25(5): 1124-1137. (in Chinese)
- [13] 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述[J]. 软件学报, 2021, 32(4): 1201-1227.
CHEN K Q, ZHU Z L, DENG X M, et al. Deep learning for multi-scale object detection: A survey[J]. *Journal of Software*, 2021, 32(4): 1201-1227. (in Chinese)
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.
- [15] 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]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 7464-7475.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]//Computer Vision - ECCV 2016. Cham: Springer, 2016: 21-37.
- [17] REN S 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.
- [18] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2980-2988.
- [19] WU T Y, DONG Y K. YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition[J]. *Applied Sciences*, 2023, 13(24): 12977.
- [20] CAO F, XING B, LUO J C, et al. An efficient object detection algorithm based on improved YOLOv5 for high-spatial-resolution remote sensing images[J]. *Remote Sensing*, 2023, 15(15): 3755.
- [21] CHEN Y P, LIU B J, YUAN L Y. PR-deformable DETR: DETR for remote sensing object detection[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 2506105.
- [22] WANG A R, XU Y, WANG H, et al. CDE-DETR: A real-time end-to-end high-resolution remote sensing object detection method based on RT-DETR[C]//IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing

Symposium. Piscataway: IEEE, 2024: 8090-8094.

- [23] ZHAO Z F, LI S Y. ABFL: Angular boundary discontinuity free loss for arbitrary oriented object detection in aerial images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5611-5611.
- [24] BAI P, XIA Y, FENG J F. Composite perception and multiscale fusion network for arbitrary-oriented object detection in remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5645-5646.
- [25] CAI X H, LAI Q X, WANG Y W, et al. Poly kernel inception network for remote sensing detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 27706-27716.
- [26] GONG M G, ZHAO H Y, WU Y, et al. Dual appearance-aware enhancement for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 62: 5602-5604.
- [27] YAO Y Q, CHENG G, LANG C B, et al. Hierarchical mask prompting and robust integrated regression for oriented object detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(12): 13071-13084.
- [28] 郭琪周, 袁春. 基于空间语义信息特征融合的目标检测与分割[J]. *软件学报*, 2023, 34(6): 2776-2788.
- GUO Q Z, YUAN C. Leveraging spatial-semantic information in object detection and segmentation[J]. *Journal of Software*, 2023, 34(6): 2776-2788. (in Chinese)
- [29] GHIASI G, LIN T Y, LE Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 7029-7038.
- [30] XU X Z, JIANG Y Q, CHEN W H, et al. Damo-yolo: A report on real-time object detection design[EB/OL]. (2022-11-23)[2025-03-17]. <https://arxiv.org/abs/2211.15444>.
- [31] XUE Y J, JU Z Y, LI Y M, et al. MAF-YOLO: Multimodal attention fusion based YOLO for pedestrian detection[J]. *Infrared Physics & Technology*, 2021, 118: 103906.
- [32] LIU C, ZHANG S X, HU M J, et al. Object detection in remote sensing images based on adaptive multi-scale feature fusion method[J]. *Remote Sensing*, 2024, 16(5): 907.
- [33] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 636-644.
- [34] LIU X M, ZHU C, YANG C Y, et al. FSFM: A feature square tower fusion module for multimodal object detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 25066-25067.
- [35] XIE X, YOU Z H, CHEN S B, et al. Feature enhancement and alignment for oriented object detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 17: 778-787.
- [36] 喻九阳, 胡天豪, 戴耀南, 等. 面向遥感目标检测的无锚框 Transformer 算法[J]. *电子学报*, 2023, 51(11): 3238-3247.
- YU J Y, HU T H, DAI Y N, et al. Anchor-free transformer algorithm for aerial remote sensing target detection[J]. *Acta Electronica Sinica*, 2023, 51(11): 3238-3247. (in Chinese)
- [37] LIU Y C, SHAO Z R, HOFFMANN N. Global Attention mechanism: Retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10) [2025-03-17]. <https://arxiv.org/abs/2112.05561>.
- [38] LLERENA J M, ZENI L F, KRISTEN L N, et al. Gaussian bounding boxes and probabilistic intersection-over-union for object detection[EB/OL]. (2021-06-10) [2025-03-17]. <https://arxiv.org/abs/2106.06072>.
- [39] LIU Z K, YUAN L, WENG L B, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. Setúbal: SCITEPRESS - Science and Technology Publications, 2017: 324-331.
- [40] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 296-307.
- [41] DIED P. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22) [2025-03-17]. <https://arxiv.org/abs/1412.6980>.
- [42] XU Y C, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(4): 1452-1459.
- [43] YANG X, YAN J C, FENG Z M, et al. R3Det: Refined single-stage detector with feature refinement for rotating object[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(4): 3163-3171.
- [44] CHENG G, WANG J B, LI K, et al. Anchor-free oriented proposal generator for object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60:

- 5625411.
- [45] HAN J M, DING J, XUE N, et al. ReDet: A rotation-equivariant detector for aerial object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 2785-2794.
- [46] DONG Y, WEI M H, GAO G S, et al. SARFA-net: Shape-aware label assignment and refined feature alignment for arbitrary-oriented object detection in remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 8865-8881.
- [47] YANG J, ZHOU L, JU Y F. AFDR-det: Adaptive feature dual-refinement oriented detector for remote sensing object detection[J]. IEEE Access, 2025, 13: 32901-32917.
- [48] LIN Q F, HUANG H B, ZHU D Y, et al. Multiple region proposal experts network for wide-scale remote sensing object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5610016.
- [49] CHENG G, YAO Y Q, LI S Y, et al. Dual-aligned oriented detector[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5618111.
- [50] ZHANG T, ZHUANG Y, CHEN H, et al. Posterior in-stance injection detector for arbitrary-oriented object detection from optical remote-sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5623918.
- [51] LIU X Y, PENG H W, ZHENG N X, et al. EfficientViT: Memory efficient vision transformer with cascaded group attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 14420-14430.
- [52] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2023: 1-5.
- [53] ZHU X Z, CHENG D Z, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 6687-6696.
- [54] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.

作者简介



曹峰 男,1980年出生于山西省朔州市.现为山西大学计算机与信息技术学院教授、博士生导师.主要研究方向为人工智能、数据挖掘与模式识别、遥感信息处理.

E-mail: caof@sxu.edu.cn



曾科文 男,2001年出生于四川省乐山市.现为山西大学计算机与信息技术学院硕士研究生.主要研究方向为高分遥感图像目标检测.

E-mail: wenClark666@163.com



李德玉 男,1965年出生于山西省临汾市.现为山西大学计算机与信息技术学院教授、博士生导师.主要研究方向为数据挖掘、计算智能、粒计算.

E-mail: lidy@sxu.edu.cn



罗喜召 男,1978年出生于河南省漯河市.现为苏州大学计算机科学与技术学院教授、博士生导师.主要研究方向为网络空间安全、人工智能安全.

E-mail: xzluo@suda.edu.cn



陶重彝 男,1985年出生于江苏省昆山市.现为苏州科技大学电子与信息工程学院副教授、硕士生导师.主要研究方向为自动驾驶态势感知、车联网安全、双足机器人运动控制、无人机目标检测.

E-mail: chongbentao@usts.edu.cn