

基于 PAM-YOLO 的遥感图像目标检测算法^{*}

梁钰漫 陈鹏宇 李 菲 王 烈 何广斌

(广西大学计算机与电子信息学院 南宁 530004)

摘 要: 针对遥感图像背景复杂、目标多尺度等特点,提出了一种基于 YOLOv8n 的遥感图像目标检测算法 PAM-YOLO。首先,在主干网络部分,构建并嵌入三池混合注意力机制(TPHA),通过融合全局最大、平均及中值池化信息,有效抑制背景噪声对通道特征的干扰,并强化关键区域的特征表达;其次,在颈部结构引入空间上下文感知模块(SCAM),通过挖掘全局上下文信息,构建通道语义与空间结构的长程依赖关系,提升模型区分目标与背景的能力;最后,设计并引入并行分支特征提取模块(PBFE),通过局部细节、全局语义和多尺度上下文 3 个并行分支的深度解耦与交互融合,实现对不同尺度目标的精细化表征。在 DIOR 数据集上,PAM-YOLO 的 mAP50 提升了 1.8%,mAP50-95 提升了 2.4%, P 值提升 0.3%, R 值提升 1.8%;在 DOTA 数据集上的各项指标也优于其他算法。实验结果表明,PAM-YOLO 在遥感图像目标检测任务中展现出更高的检测精度与鲁棒性。

关键词: 遥感图像;目标检测;YOLOv8n;注意力机制;多尺度特征融合

中图分类号: TP919.5;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.2060

PAM-YOLO for object detection in remote sensing images

Liang Yuman Chen Pengyu Li Fei Wang Lie He Guangbin

(School of Computer and Electronic Information, Guangxi University, Nanning 530004, China)

Abstract: To address the characteristics of complex backgrounds and multi-scale objects in remote sensing images, a remote sensing image object detection algorithm based on YOLOv8n, named PAM-YOLO, is proposed. First, in the backbone network, a triple pool hybrid attention (TPHA) mechanism is constructed and embedded. By fusing information from global maximum, average, and median pooling, it effectively suppresses the interference of background noise on channel features and enhances the feature representation of key regions. Second, a spatial context aware module (SCAM) is introduced in the neck structure. By mining global contextual information and building long-range dependencies between channel semantics and spatial structure, it improves the model's ability to distinguish between targets and backgrounds. Finally, a parallel branch feature extraction (PBFE) module is designed and introduced. Through the deep decoupling and interactive fusion of three parallel branches for local details, global semantics and multi-scale context, it achieves a refined representation for objects of different scales. On the DIOR dataset, PAM-YOLO's mAP50 increased by 1.8%, mAP50-95 by 2.4%, P -value by 0.3%, and R -value by 1.8%; its metrics on the DOTA dataset are also superior to other algorithms. The experimental results indicate that PAM-YOLO shows higher detection accuracy and robustness in the task of remote sensing image object detection.

Keywords: remote sensing image; object detection; YOLOv8n; attention mechanism; multi-scale feature fusion

0 引 言

近几十年来,我国航空航天技术持续迅猛发展,获取的遥感图像数量与质量均大幅提升,这使得遥感图像的研究成为热门课题。在众多相关研究中,遥感图像目标检测技术备受瞩目^[1-4]。作为遥感图像研究领域的关键任务,遥感

图像目标检测的性能优劣,直接影响着船舶动态监测、交通秩序监测与管理、城市建筑物规划等工作的开展^[5-6]。

早期的传统检测算法主要依赖滑动窗口与人工设计特征,不仅计算冗余、效率低下,且在面对复杂多变的遥感场景时鲁棒性较差,难以满足实际应用需求。近年来,卷积神经网络的兴起极大地推动了该领域的发展,尤其是 YOLO

(you only look once)系列算法^[7],凭借其高效的端到端检测能力,已成为遥感目标检测的主流框架。然而,遥感图像普遍存在背景纹理复杂、目标多尺度等特点,这对检测算法的特征表征鲁棒性与尺度适应性提出了严峻挑战。针对这些问题,国内外的研究人员主要从模型轻量化设计和特征增强与多尺度融合两个维度进行了广泛探索。

第1类研究主要致力于模型的轻量化与结构精简。为此,研究者们通常采用网络剪枝、结构重组或引入轻量化卷积等策略。例如,LAI-YOLOv5s^[8]通过替换检测头并引入轻量级特征融合模块(DFM-CPFN),旨在降低模型空间复杂度;GUAV-YOLO^[9]和CM-YOLOv8s^[10]则通过去除大目标检测头、重构颈部结构等方式,显著降低了模型参数量;基于YOLOv8改进的RSD-YOLO^[11]采用LAMP剪枝技术并结合轻量化主干,实现了更为紧凑的模型表征;改进RT-DETR^[12]引入部分卷积(PConv)构建基础模块,在保证基础特征表达的同时优化了参数利用率。尽管这些轻量化方法在降低模型规模方面效果显著,但大幅度的参数削减往往伴随着深层语义信息的丢失,导致模型在面对背景干扰或特征表现较弱的目标时特征提取能力受限,难以满足高精度检测的需求。

第2类研究则侧重于特征提取能力的增强与多尺度感知,通常采用注意力机制、改进特征金字塔(feature pyramid network, FPN)等方法。基于YOLOv8改进的YOLO-SE^[13]和FEW-YOLOv8^[14]通过引入EMA注意力机制或重构C2f模块,增强了网络对多尺度目标的敏感度;CSYOLOv7^[15]和改进YOLOv11^[16]设计了跨尺度特征提取模块和高效的多尺度FPN结构,以优化特征的尺度适应性;针对复杂背景,DT-YOLOv8^[17]通过嵌入BiFormer注意力机制并设计特征交互融合的DAT检测头,显著增强了模型对目标的关注与特征表达能力;而改进的Faster-RCNN^[18]和YOLOv8-LGA^[19]分别利用显著性引导模块和大选择核注意力来捕捉空间信息。这些方法虽然在一定程度上提升了检测精度,但仍存在局限性:许多注意力机制在聚合通道信息时,其依赖的全局池化策略在复杂的遥感背景下存在固有冲突,难以同时抑制噪声并凸显目标;此外,现有的特征融合策略在构建长程依赖和解耦多尺度语义信息上仍显不足,导致模型难以充分适应遥感图像中目标多尺度的特性。

鉴于此,为了弥补轻量级模型在复杂遥感场景下特征提取能力不足的缺陷,并在检测精度与模型复杂度之间寻求更优平衡,本文基于轻量级YOLOv8n提出了一种改进的遥感图像目标检测算法——PAM-YOLO。本文在将计算开销控制在合理范围内的前提下,将优化重心聚焦于提升模型对复杂环境的抗干扰能力与多尺度目标的表征能力。通过在DIOR和DOTA数据集上的实验,验证了该算法的有效性。本研究的主要工作与贡献为:

1)针对遥感图像背景复杂导致的噪声干扰问题,构建

三池混合注意力机制(triple pool hybrid attention, TPHa),通过融合全局最大、平均及中值池化信息,有效抑制背景噪声对通道特征的干扰,并强化关键区域的特征表达;

2)针对复杂背景中目标与背景的特征混淆问题,引入空间上下文感知模块(spatial context aware module, SCAM),通过挖掘全局上下文信息,构建通道语义与空间结构的长程依赖关系,从而提升模型区分目标与背景的能力;

3)针对遥感场景中目标尺度多样性问题,设计并行分支特征提取模块(parallel branch feature extraction, PBFE),通过局部分支、全局分支及多尺度分支的并行交互,深度解耦特征、实现跨尺度融合,实现对微小目标到大型目标的精细化表征,提升多尺度目标检测鲁棒性与表征能力。

1 YOLOv8 目标检测算法

YOLOv8算法是Ultralytics公司2023年研发的一种单阶段检测算法,已在目标检测、图像分类等领域获得广泛应用。YOLOv8的整体网络结构可分为主干网络、颈部结构以及头部结构3个部分。

主干网络以CSP Darknet为基础架构,整合了C2f和SPPF等关键模块。其中,C2f模块通过引入交叉卷积架构,结合Bottleneck Block与SPPF模块,在维持较低参数量的同时,削减了冗余参数,提升了计算效率。SPPF模块则通过多尺度池化操作实现特征提取与整合,增强了模型对不同尺寸目标的识别能力,使模型能够灵活应对目标尺寸的变化,从而进一步提升检测精度和鲁棒性。

颈部结构采用了FPN模块和PANet模块。FPN将深层语义特征向浅层传递,提高了语义信息的表达能力,但未实现浅层定位信息的向上传递。PANet在FPN的基础上进行了扩展,通过引入自下而上的金字塔路径,使浅层的高精度定位特征能够向上传播,实现了深层与浅层特征的融合。

在头部部分,YOLOv8采用了Anchor-Free检测头设计,减少了对预设锚框的依赖,提升了模型的灵活性和泛化能力。在损失函数的选择上,YOLOv8使用VFL作为分类任务的损失函数,并使用DFL结合CIoU损失函数来处理边界框回归任务。这种组合能够更精准地量化模型预测结果与真实标签之间的偏差,提升模型的拟合精度。

2 PAM-YOLO 算法

PAM-YOLO网络结构如图1所示。PAM-YOLO主要在主干网络和颈部结构进行结构优化。首先在主干网络部分,构建并嵌入TPHa,通过融合全局最大、平均及中值池化信息,有效抑制背景噪声对通道特征的干扰,并强化关键区域的特征表达;其次,在颈部结构引入SCAM,通过挖掘全局上下文信息,构建通道语义与空间结构的长程依赖

关系,提升模型区分目标与背景的能力;最后,设计并引入 PBFE,通过局部细节、全局语义和多尺度上下文 3 个并行

分支的深度解耦与交互融合,实现对不同尺度目标的精细化表征。

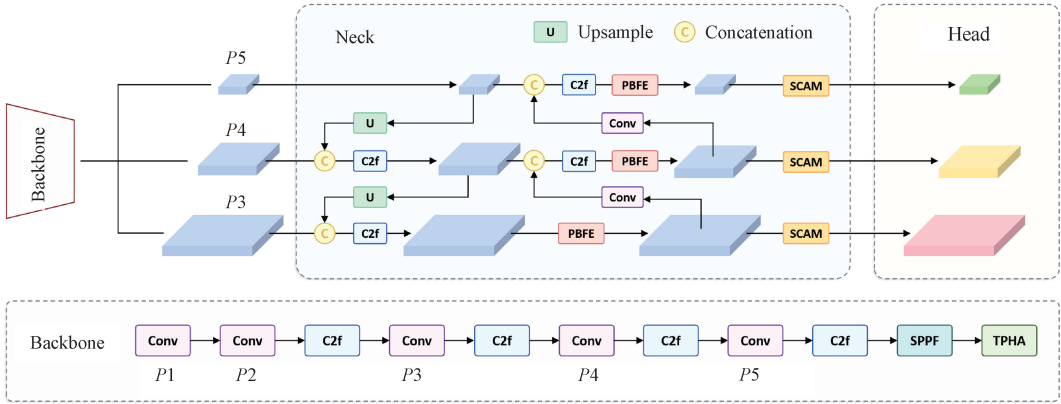


图 1 PAM-YOLO 网络结构
Fig. 1 PAM-YOLO network structure

2.1 TPHA

遥感图像固有的广覆盖、富信息特性,使其背景成分异常复杂。这种复杂背景引入的噪声严重干扰了目标物体的准确识别,极大影响下游任务的性能。

当前主流的通道注意力机制通常依赖全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling, GMP)来提取特征图的全局统计信息。然而,

这两种方式各有局限:GAP 对整体特征进行编码,容易因背景信息而稀释目标信号;GMP 能强化局部显著响应,但对异常值或噪声点敏感。尽管已有研究尝试将二者结合,但这种简单的策略未能化解各自的缺陷,导致模型依然难以在抑制背景噪声的同时,有效保留关键的目标特征。

为解决这一问题,本文构建并引入了一种 TPHA^[20],其结构如图 2 所示。

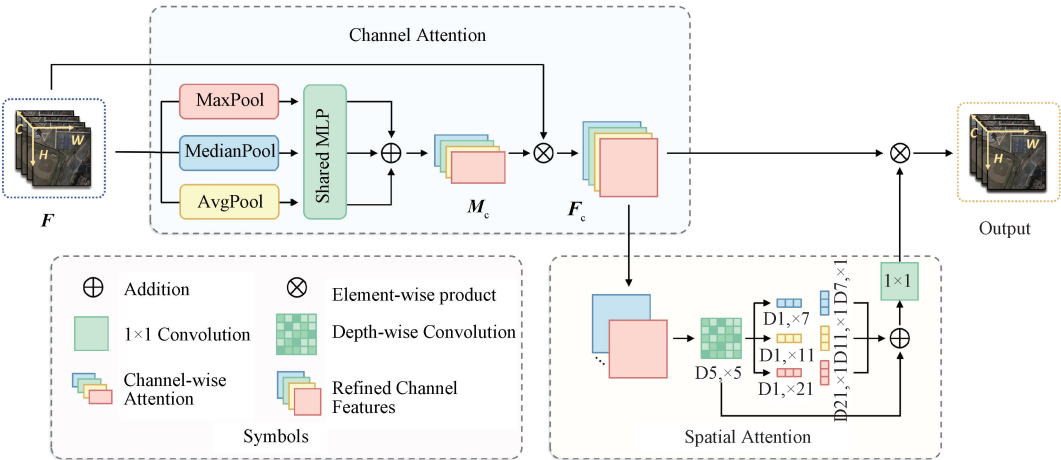


图 2 TPHA 结构
Fig. 2 Structure of TPHA

TPHA 在通道维度上采用 3 种池化操作。它使用平均池化(AvgPool)与最大池化(MaxPool)来捕捉特征的全局平均分布与峰值响应。作为关键补充的中值池化(MedianPool)凭借其对离群值不敏感的统计特性,提供了一个稳定的全局基准,能够缓解背景噪声对平均池化的干扰以及异常值对最大池化的影响。这 3 个池化操作分别生成特征向量为 $F_{\max}$ 、 F_{median} 、 F_{avg} 。随后,其被送入一个共享权重的多层感知机(multi-layer perceptron, MLP)中,学习通道间的非线性依赖关系。最终,通过元素求和与

Sigmoid 激活函数融合三者输出,生成最终的通道注意力权重 M_c :

$$M_c = \sigma(\text{MLP}(F_{\max}) + \text{MLP}(F_{\text{median}}) + \text{MLP}(F_{\text{avg}})) \quad (1)$$

式中: σ 表示 sigmoid 函数。该权重 M_c 通过广播机制与原始输入特征图 F 逐通道相乘,增强目标相关通道并抑制噪声主导通道,得到具备更强抗噪声的加权特征图 F_c 。

接着,特征图 F_c 进入空间注意力模块(spatial attention)。模块首先采用 5×5 深度卷积提取基础空间特

征,以保留更多细节。为了在不显著增加计算成本的前提下扩大感受野,使用了一组包含 6 个不同尺寸的条状深度卷积核($7 \times 1, 1 \times 7, 11 \times 1, 1 \times 11, 21 \times 1, 1 \times 21$)。这种设计通过模拟大尺寸卷积核的效果,捕获了多方向、多尺度的上下文信息,显著增强了模型对复杂空间结构的感知能力,该过程如式(2)所示。

$$\mathbf{F}_s = \sum_{k \in \{7, 11, 21\}} (\text{DWConv}_{k \times 1} (\text{DWConv}_{5 \times 5} (\mathbf{F}_c))) + \text{DWConv}_{1 \times k} (\text{DWConv}_{5 \times 5} (\mathbf{F}_c)) \quad (2)$$

随后,通过一个 1×1 卷积对上述多尺度特征进行通道集成与信息混合,并经由 Sigmoid 函数归一化,生成最终的空间注意力权重 $\mathbf{M}_s$:

$$\mathbf{M}_s = \sigma (\text{Conv}_{1 \times 1} (\mathbf{F}_s)) \quad (3)$$

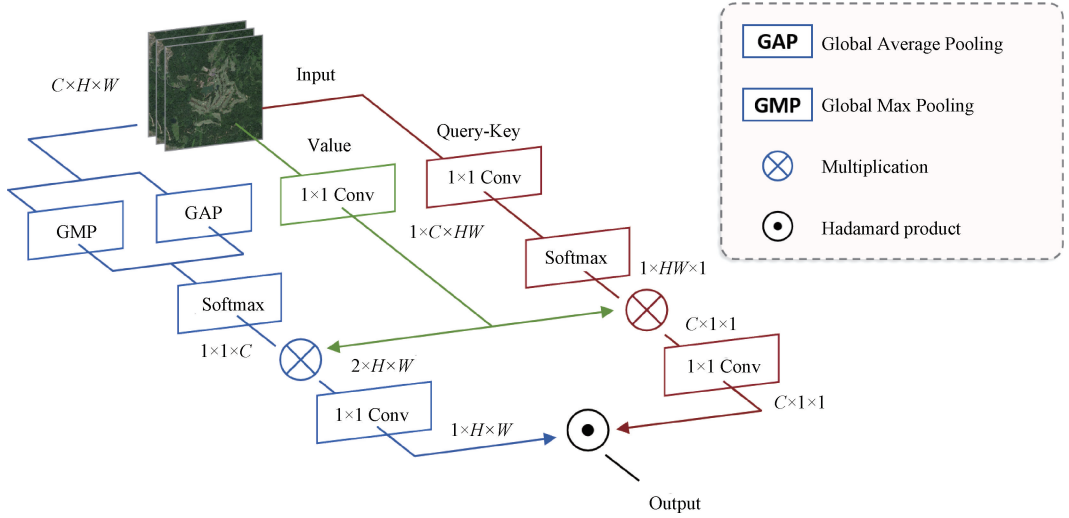


图 3 SCAM 结构

Fig. 3 Structure of SCAM

设输入特征图为 $\mathbf{X}$, 其维度为 $C \times H \times W$, 其中 C 为通道数, H 和 W 为空间尺寸。

第 1 个分支采用 GAP 和 GMP 分别提取输入特征图在通道维度上的平均及最大值信息。所得结果经 Softmax 归一化后,生成通道维度的权重 ($1 \times 1 \times C$),实现对通道维度的动态加权,从而增强包含关键信息的通道响应,抑制冗余特征,在通道层面提升目标与背景的语义区分度。

第 2 个分支通过一个 1×1 卷积对输入特征 $\mathbf{X}$ 进行线性变换,生成特征表示 Value。此操作保持通道数不变,并将调整形状为 ($1 \times C \times HW$),为后续的注意力计算提供特征基础空间。

第 3 个分支同样采用 1×1 卷积以生成 Query-Key 对。随后,在空间维度上进行 Softmax 归一化,生成空间注意力权重 ($1 \times HW \times 1$)。该权重矩阵旨在显式地建模像素间的长程依赖关系,捕捉目标与背景之间的全局空间关联性。

在特征融合阶段,模块将第 1 分支的通道权重与第 2 分支的 Value 进行矩阵乘法,强化了特征的通道语义。同

式中: σ 表示 sigmoid 函数。最终,空间权重 $\mathbf{M}_s$ 与经过通道加权的特征图 $\mathbf{F}$ 进行逐元素相乘,得到 TPHA 模块的最终输出。

2.2 SCAM

在遥感图像中,目标与背景在光谱、纹理或几何特征上的高度相似性,常常导致目标的判别性特征被复杂背景淹没,进而影响算法区分目标与背景区域的能力,易引发误检与漏检。现有的特征提取方法在捕捉局部特征方面表现出色,但对全局空间关联性的建模能力有限,难以有效利用上下文信息来区分相似物体。

为增强特征的判别力,弥补这一不足,本文引入了 SCAM^[21]。SCAM 的结构如图 3 所示,其通过 3 个分支协同工作,以构建更具判别力的特征表示。

时,将第 3 分支的空间权重与第 2 分支的 Value 进行矩阵乘法,聚合了跨空间的上下文信息。最后,这两类经过上下文增强的特征通过哈达玛乘积进行深度融合。这一过程构建了通道语义与空间结构的联合表示,使模型能够在复杂遥感场景下更有效地区分目标与背景,从而显著增强特征的判别能力。

2.3 并行分支特征提取 PBFE

遥感图像中往往同时存在数量众多、尺度跨度极大的不同类别目标,如机场、港口等大尺度目标与车辆、船舶等小尺度目标。这种大范围、多目标、多尺度分布的特性对检测模型的多尺度特征感知和表达能力提出了更高要求。

Lu 等^[22]提出了一种将特征流并行解耦为“局部流”与“全局流”的通用思想,该双分支虽能有效兼顾细节与语义,但其并未对遥感图像固有的尺度差异问题进行专门优化。为此,本文受此启发并结合遥感图像中目标多尺度的特性,设计了多尺度分支,从而提出了一种并行分支特征提取模块 PBFE,其结构如图 4 所示。

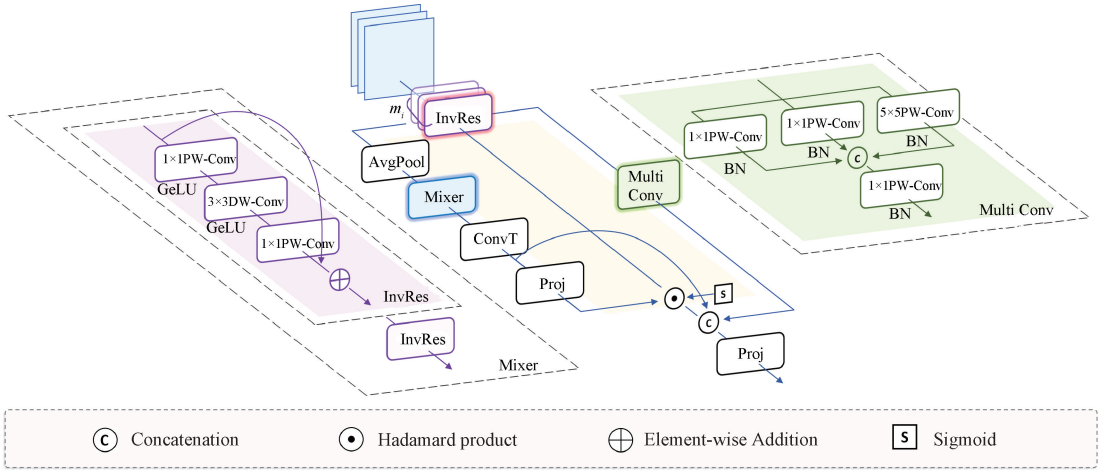


图 4 PBFE 结构

Fig. 4 Structure of PBFE

该模块采用并行 3 分支结构,针对输入特征进行深度解耦和跨尺度交互,利用局部分支保留细节特征、全局分支捕获语义信息,并通过多尺度分支实现跨尺度特征融合,从而能够支持从微小目标到大型目标的特征表达,提升多尺度目标检测的鲁棒性与表征能力。

具体而言,PBFE 首先利用 m_i 个连续的倒置残差块 (InvRes) 对输入特征图 X_i^l 进行预处理。每个 InvRes 块包含 1×1 卷积、GeLU 激活函数和 3×3 深度卷积。在该模块中,先通过 1×1 的卷积来扩展通道数,捕获复杂特征并减少细节信息损失。随后对扩展后的特征应用 GeLU 激活函数,增强非线性表达能力。接着,使用 3×3 卷积核,按通道分组进行深度卷积,捕捉边缘、纹理等局部细节,深度卷积后再次引入非线性。然后,使用 1×1 卷积将通道数压缩回原始输入通道数,既保留关键信息又降低计算量。最后将模块的输入与处理后的输出逐元素相加,这一过程即残差连接,能够缓解梯度消失与特征退化问题。

经过 m_i 个 InvRes 块处理后,特征图 X_i^l 被送入 3 个并行分支:局部特征提取分支、全局特征提取分支和多尺度特征融合分支。

局部特征提取分支保持原始特征图的大小,不加注意力处理或下采样操作,直接将 X_i^l 传递到块的末尾。其核心目的是无损地保留图像的原始细节,特别是边缘、纹理等高频信息,这对于目标的精确检测至关重要。

全局特征提取分支中,首先采用平均池化对特征图进行下采样。随后,通过两层连续的 InvRes 块 (Mixer),以强化全局语义信息的表征能力。最后,利用转置卷积 (ConvT) 恢复原始空间分辨率,生成富含全局上下文的特征图 X_i^g ,如式(4)所示。

$$X_i^g = \text{ConvT}(\text{Mixer}(\text{AvgPool}(X_i^l))) \quad (4)$$

多尺度特征融合分支采用不同感受野的卷积核并行处理输入特征,以捕获多样的尺度信息。提取出的多尺度特征在通道维度上进行拼接,以集成不同粒度的上下文信

息。随后,通过 1×1 卷积压缩通道数,输出具有丰富尺度信息的特征图 X_i^m ,如式(5)所示。

$$X_i^m = \text{Conv}_{1 \times 1} \left(\text{Conv}_{1 \times 1}(X_i^l) \parallel \text{Conv}_{3 \times 3}(X_i^l) \parallel \text{Conv}_{5 \times 5}(X_i^l) \right) \quad (5)$$

在分支特征融合阶段,首先对全局分支的输出特征图 X_i^g 进行逐点卷积与批量归一化 (Proj) 生成中间特征,经 Sigmoid 函数映射至 $[0, 1]$ 区间,得到权重图 W_i^g ,如式(6)所示。

$$W_i^g = \sigma(\text{Proj}(X_i^g)) \quad (6)$$

式中: σ 表示 sigmoid 函数。

接着,将权重图 W_i^g 与局部特征图 X_i^l 通过哈达玛乘积加权融合,实现由全局语义引导的局部细节优化。最后,将加权后的局部特征 $X_i^l \odot W_i^g$ 与全局特征 X_i^g 和多尺度特征 X_i^m 在通道维度拼接,构建了一个信息维度更丰富、表达能力更强的统一特征空间 X_i^u ,如式(7)所示。

$$X_i^u = (X_i^l \odot W_i^g \parallel X_i^g \parallel X_i^m) \quad (7)$$

式中: $\odot$ 表示哈达玛乘积。PBFE 模块末尾通过投影层将输出通道数减半,以控制模型复杂度,最终得到模块的输出特征图。

3 实验结果与分析

3.1 实验环境和数据集

实验在 Windows10 系统上运行, GPU 为 NVIDIA GeForce RTX 3080 Ti, 采用的深度学习框架为 PyTorch 2.0.0, CUDA 版本为 11.7, 编译语言为 Python 3.8, YOLOv8n 作为基准网络模型, 输入图片大小为 640×640 , 优化器为 SGD, 训练次数为 300, 批量大小为 8。实验所用数据集均按 7:2:1 的比例划分为训练集、验证集和测试集。

实验采用 DIOR 数据集^[23]和 DOTA 数据集^[24]。西北工业大学程焱团队提出的 DIOR 数据集是一个光学遥感

图像目标检测的大规模基准数据集,其图像数据来源于 Google Earth、高分二号卫星、吉林一号卫星和 Cyclo Media B. V. 航拍影像。该数据集规模较大,共包含 23 463 幅图像,涵盖 20 个典型目标类别,这些类别包括飞机、机场、棒球场、篮球场、桥梁、烟囱、水坝、高速公路服务区、高速公路收费站、港口、高尔夫球场、地面田径场、天桥、船舶、体育场、储罐、网球场、火车站、车辆和风磨等。

DOTA 数据集由武汉大学提出,该数据集由 2 806 张来自不同传感器和平台的多分辨率航空图像组成,包含 15 个不同类别的对象,包括飞机、船舶、储罐、棒球场、网球场、篮球场、地面田径场、港口、桥梁、大型车辆、小型车辆、直升机、环形交叉路口、足球场和游泳池。该数据集总计包含 188 282 个目标实例,这些实例均通过面向对象的包围框进行标注。每张图像中均存在多个目标,且这些目标在尺度、方向和形状上均有所不同。

3.2 评价指标

本文将准确率 P 、召回率 R 、平均精度均值 $mAP50$ 和 $mAP50-95$ 作为评价指标,来衡量算法的检测效果。

准确率 P 和召回率 R 的计算公式分别如式(8)、(9)所示。

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

其中,TP 代表正例被识别为正例的图像数量,FP 代表负例被识别为正例的图像数量,FN 代表正例被识别为负例的图像数量。

精度 AP 和平均精度均值 mAP 的计算公式分别如式(10)、(11)所示。

$$AP = \int_0^1 P(R) dR \tag{10}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{11}$$

其中, N 是类别数量, AP_i 为第 i 个类别的平均精度。 $mAP50$ 是所有类别的 IoU 阈值在 0.5 时的平均检测精度; $mAP50-95$ 表示所有类别的 IoU 阈值在 0.5~0.95 间的平均检测精度,步长为 0.05。

3.3 消融实验

为验证全局中值池化在 TPHA 模块中的必要性,在 DIOR 数据集上进行了实验,实验结果如表 1 所示。

表 1 TPHA 模块消融实验结果
Table 1 TPHA module ablation experiment results

模型	/%			
	mAP50	mAP50-95	P	R
Model1	72.0	49.9	82.2	65.3
Model2	76.6	54.4	85.5	68.6

注:Model1 表示在 TPHA 中未引入全局中值池化,Model2 表示在 TPHA 中引入了全局中值池化。

由表 1 可知,引入全局中值池化的 TPHA 模块(Model2)在所有检测指标上均大幅提升。与未引入全局中值池化的 Model1 相比,Model2 的 $mAP50$ 和 $mAP50-95$ 分别提升 4.6% 和 4.5%, P 和 R 分别提高 3.3% 和 3.3%。该结果表明,全局中值池化作为对全局平均池化和最大池化的有益补充,校正了平均池化因背景信息而稀释目标信号以及最大池化易受噪声或异常值干扰的固有缺陷。通过这种互补,TPHA 能够生成更可靠的通道权重,精确地增强目标相关特征并抑制背景噪声。

为了验证本文所使用模块的有效性,以 YOLOv8n 为基准模型,对不同的改进方法在 DIOR 数据集上进行消融实验,实验结果如表 2 所示。

表 2 消融实验结果
Table 2 Ablation experiment results

YOLOv8n	TPHA	SCAM	PBFE	mAP50	mAP50-95	P	R
✓	×	×	×	76.5	54.1	85.7	68.7
✓	✓	×	×	76.6	54.4	85.5	68.6
✓	✓	✓	×	77.1	54.4	85.9	69.6
✓	✓	×	✓	77.5	55.7	85.5	70.2
✓	×	✓	✓	78.2	56.2	85.3	70.5
✓	✓	✓	✓	78.3(+1.8)	56.5(+2.4)	86.0(+0.3)	70.5(+1.8)

注:“×”表示未引用该模块;“✓”表示引用该模块。

实验结果显示,仅在 YOLOv8n 基础上引入 TPHA 模块后, $mAP50$ 和 $mAP50-95$ 分别提升了 0.1% 和 0.3%,表明 TPHA 模块通过抑制噪声和增强目标关键结构特征,降低了噪声引发的误检和漏检,尽管 P 和 R 略有波动,这是对低置信度样本进一步筛选所致,是特征增强与噪声抑

制之间权衡的合理代价。

在双模块组合的实验中,引入 TPHA 和 SCAM 模块后, $mAP50$ 较基线提升 0.6%,说明 SCAM 模块通过建模空间长程依赖,优化了目标与背景的区分能力。而当采用 TPHA 与 PBFE 模块组合时, $mAP50$ 提升至 77.5%,

mAP50-95 提升至 55.7%，这一显著增长表明 PBFE 的并行分支结构能有效捕捉多尺度特征，弥补了单一尺度特征提取的不足。此外，实验还测试了 SCAM 与 PBFE 模块的组合(不含 TPHA)，其 mAP50 达到了 78.2%，这进一步验证了全局上下文感知与多尺度特征解耦在复杂遥感场景下的核心作用。

当 3 个模块同时应用时，mAP50 最终达到 78.3%。虽然该结果相较于“SCAM+PBFE”组合 mAP 上表现接近，但准确率 P 却显著提升了 0.7%。这说明，在 PBFE 和 SCAM 提取丰富特征的基础上，TPHA 进一步过滤了背景

噪声，有效减少了误检。最终，三者协同工作，在准确率 P 、召回率 R 及 mAP 各项指标上均实现了最佳平衡。

3.4 对比实验

为进一步验证 PAM-YOLO 算法的有效性与先进性，本文在 DIOR 数据集上与现有的主流目标检测算法进行了对比，结果如表 3 所示。实验选取的对比模型涵盖了从经典的 YOLOv5s、YOLOv8n 到最新的 YOLOv10n、YOLOv11n，同时也包括了 CM-YOLOv8s、改进的 Faster-RCNN、CSYOLOv7 等针对遥感图像特点进行特定优化的最新改进算法。

表 3 DIOR 数据集各算法对比实验结果

算法	mAP50/ %	mAP50-95/ %	P / %	R / %	参数量/ M	GFLOPs
YOLOv5n	76.0	53.0	84.8	67.8	2.19	5.8
YOLOv6n	74.8	53.7	86.2	66.3	4.16	11.5
YOLOv8n	76.5	54.1	85.7	68.7	2.68	6.8
YOLOv9t	77.5	56.2	86.6	69.6	1.73	6.5
YOLOv10n	77.0	54.9	86.7	68.1	2.70	8.3
YOLOv11n	77.9	55.7	86.6	69.7	2.59	6.3
CM-YOLOv8s ^[10]	70.9	41.8	75.4	67.6	—	—
CSYOLOv7 ^[15]	74.0	—	83.1	68.3	7.40	19.1
改进 Faster-RCNN ^[18]	69.8	49.3	76.6	65.3	—	—
PAM-YOLO	78.3	56.5	86.0	70.5	4.55	10.3

横向对比各模型指标可知，PAM-YOLO 在各项关键性能指标上均实现了显著提升。具体而言，PAM-YOLO 的 mAP50 达到了 78.3%，mAP50-95 达到 56.5%，召回率 R 达到 70.5%，均优于参与对比的其他算法。该数据充分说明，相较于现有主流算法，PAM-YOLO 在应对遥感图像复杂背景及多尺度目标方面展现出了更优越的综合检测能力。

3.5 泛化实验

为评估 PAM-YOLO 算法的泛化能力，本文在 DOTA 数据集上进行了详细对比，结果如表 4 所示。实验选取了包括 RSD-YOLO、FEW-YOLOv8、DT-YOLOv8 及 YOLOv8-LGA 在内的多种基于 YOLOv8 改进的同类前沿算法作为对比对象。

表 4 DOTA 数据集各算法对比实验结果

算法	mAP50/ %	mAP50-95/ %	P / %	R / %	参数量/ M	GFLOPs
YOLOv5n	76.2	52.8	80.3	71.6	2.19	5.8
YOLOv6n	73.0	49.6	80.6	67.9	4.16	11.5
YOLOv8n	77.9	54.7	81.9	72.8	2.68	6.8
RSD-YOLO ^[11]	73.0	50.4	76.6	68.7	8.27	42.7
改进 RT-DETR 的算法 ^[12]	72.3	48.5	77.6	69.6	17.30	—
FEW-YOLOv8 ^[14]	69.3	—	73.5	66.4	11.19	29.5
改进 YOLOv11 的算法 ^[16]	72.5	47.9	81.7	70.5	5.80	11.8
DT-YOLOv8 ^[17]	73.5	49.4	78.6	70.3	5.40	—
YOLOv8-LGA ^[19]	73.6	—	81.6	68.8	—	—
PAM-YOLO	80.5	57.7	83.7	74.1	4.55	10.3

可以看出，PAM-YOLO 在 DOTA 数据集上依然保持了最佳的检测精度，其 mAP50、mAP50-95、 P 和 R 指

标相较于 YOLOv8n 分别提升 2.6%、3.0%、1.8% 和 1.3%。上述结果进一步证明了 PAM-YOLO 在遥感图像

目标检测任务具有出色的泛化能力, 并能够保持良好的检测精度和鲁棒性。此外, 虽然 PAM-YOLO 的参数量相较于基准模型 YOLOv8n 有所增加, 但该增加量仍低于同类改进算法如 RSD-YOLO 和 FEW-YOLOv8, 且换取了显著的精度提升, 在检测精度与模型复杂度之间取得了更优的平衡。

4 检测结果可视化

为直观地展示所提出算法在遥感图像目标检测任务中的性能, 分别使用 YOLOv8n 和 PAM-YOLO 算法在 DIOR 和 DOTA 数据集上进行可视化检测, 检测结果如图 5 和 6 所示。

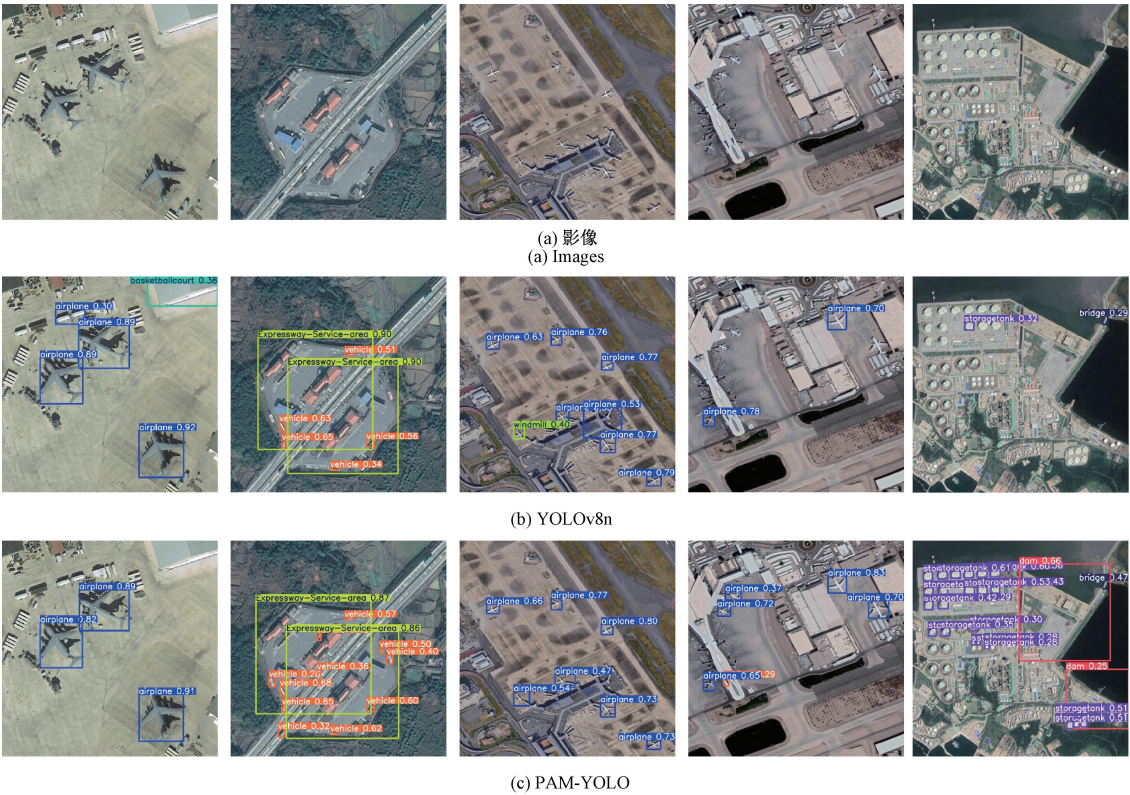
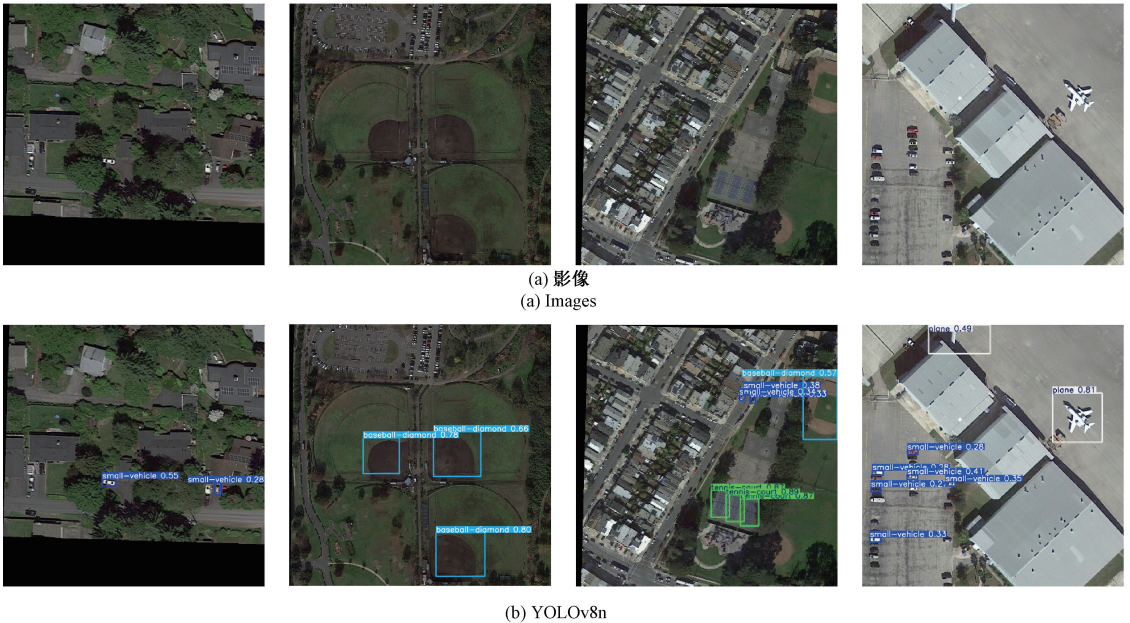
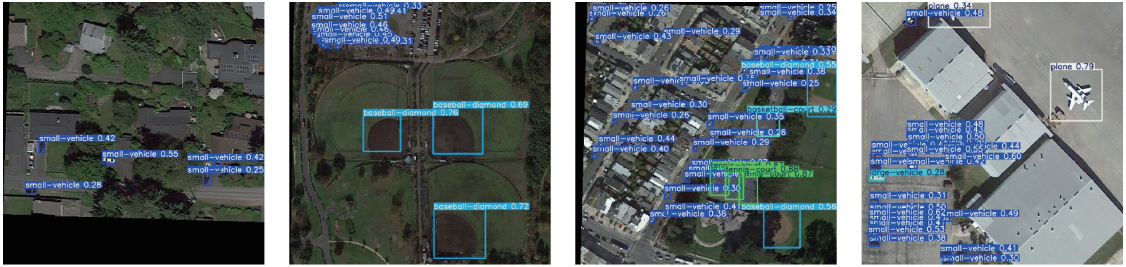


图 5 在 DIOR 数据集上的检测效果
Fig. 5 Detection performance on the DIOR dataset





(c) PAM-YOLO

图 6 在 DOTA 数据集上的检测效果

Fig. 6 Detection performance on the DOTA dataset

由图 5 和 6 可知, YOLOv8n 在面对复杂遥感场景时存在明显局限性, 主要表现为难以有效区分目标与相似背景区域, 导致部分目标漏检或误检, 尤其是在目标密集、尺度变化大或背景杂乱的情况下更为突出。相比之下, PAM-YOLO 能够更加准确地定位并识别图像中的各类目标, 显著减少误检和漏检现象。

综上, 通过可视化结果直观验证了 PAM-YOLO 在遥感图像目标检测任务中的优越性。该算法不仅提升了目标检测的准确率与鲁棒性, 同时在复杂背景和多尺度目标环境下展现出更高的实际应用价值。

5 结 论

针对遥感图像中背景复杂、目标多尺度等特点, 基于 YOLOv8n 提出一种遥感图像目标检测算法 PAM-YOLO。在主干网络部分, 构建并嵌入 TPHa, 通过融合全局最大、平均及中值池化信息, 有效抑制背景噪声对通道特征的干扰, 并强化关键区域的特征表达; 其次, 在颈部结构引入 SCAM, 通过挖掘全局上下文信息, 构建通道语义与空间结构的长程依赖关系, 提升模型区分目标与背景的能力; 最后, 设计并引入 PBF, 通过局部细节、全局语义和多尺度上下文 3 个并行分支的深度解耦与交互融合, 实现对不同尺度目标的精细化表征。

在 DIOR 和 DOTA 数据集上的实验结果表明, 所提 PAM-YOLO 算法在准确率 P 、召回率 R 、 mAP_{50} 及 mAP_{50-95} 等评价指标上均表现出色, 展现出良好的目标检测性能与泛化能力。在 DIOR 数据集上, PAM-YOLO 相较于基准模型 YOLOv8n, 其 mAP_{50} 提升了 1.8%, mAP_{50-95} 提升了 2.4%。为进一步验证其泛化能力, 在 DOTA 数据集上的对比实验中, 本算法的 mAP_{50} 和 mAP_{50-95} 更是取得了 2.6% 和 3.0% 的显著提升, 并在所有对比算法中均展现出最佳性能。可视化结果也直观证明, 本算法能有效减少漏检与误检。

参考文献

[1] 孙晓永, 孙备, 郭润泽, 等. 面向多无人机协同的多模态目标检测方法[J]. 仪器仪表学报, 2025, 46(2): 209-220.

SUN X Y, SUN B, GUO R Z, et al. Multimodal target detection method for multi-UAV coordination[J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 209-220.

[2] 闫钧华, 张琨, 施天俊, 等. 融合多层级特征的遥感图像地面弱小目标检测[J]. 仪器仪表学报, 2022, 43(3): 221-229.

YAN J H, ZHANG K, SHI T J, et al. Multi-level feature fusion based dim small ground target detection in remote sensing images[J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.

[3] 冉宁, 施高朗, 张少康, 等. 基于 YOLOv8 的遥感小目标检测算法[J]. 电子测量与仪器学报, 2025, 39(5): 197-207.

RAN N, SHI G L, ZHANG S K, et al. Remote sensing small target detection algorithm based on YOLOv8[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 197-207.

[4] 王海群, 武泽锴, 晁帅. 改进 YOLOv8n 的遥感图像目标检测算法[J]. 电子测量与仪器学报, 2025, 39(4): 84-94.

WANG H Q, WU Z K, CHAO S. Improve the YOLOv8n object detection algorithm for remote sensing images[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(4): 84-94.

[5] 魏庭旭, 陈颖, 李铖昊, 等. 融合双域特征与跨维门控注意力的遥感图像配准[J]. 电子测量技术, 2025, 49(4): 217-226.

WEI T X, CHEN Y, LI C H, et al. Remote sensing image registration fusing dual-domain features and cross-dimensional gated attention [J]. Electronic Measurement Technology, 2025, 49(4): 217-226.

[6] 雷帮军, 余翱, 吴正平, 等. 基于位置感知和跨层特征融合的航拍小目标检测算法[J]. 电子测量技术, 2024, 47(5): 112-123.

- LEI B J, YU A, WU Z P, et al. Small target detection algorithm for aerial photography based on location awareness and cross-layer feature fusion[J]. Electronic Measurement Technology, 2024, 47(5): 112-123.
- [7] 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.
- [8] DENG L X, BI L Y, LI H Q, et al. Lightweight aerial image object detection algorithm based on improved YOLOv5s[J]. Scientific Reports, 2023, 13(1): 7817.
- [9] 高仪, 陈鹏, 刘子奇, 等. GUAV-YOLO:面向灰度影像无人机小目标的轻量化检测模型[J]. 光学学报, 2025, 45(23): 98-113.
- GAO Y, CHEN P, LIU Z Q, et al. GUAV-YOLO: Lightweight detection model for UAV small targets in grayscale images [J]. Acta Optica Sinica, 2025, 45(23): 98-113.
- [10] 廖宁生, 曹天秀, 刘科言, 等. 复合特征与多尺度融合的无人机小目标检测算法[J]. 计算机工程与应用, 2025, 61(3): 111-120.
- LIAO N S, CAO T X, LIU K Y, et al. Small target detection algorithm for UAV based on composite feature and multi-scale fusion[J]. Computer Engineering and Applications, 2025, 61(3): 111-120.
- [11] 汤伟博, 方强, 李沛根, 等. 基于RSD-YOLO的无人机航拍图像的小目标检测[J]. 计算机工程, 2026, 52(4): 214-228.
- TANG W B, FANG Q, LI P G, et al. Small target detection in UAV aerial images based on RSD-YOLO[J]. Computer Engineering, 2026, 52(4): 214-228.
- [12] 刘思元, 高凯, 雍龙泉. 改进 RT-DETR 的航拍小目标检测算法[J]. 计算机工程与应用, 2025, 61(4): 272-281.
- LIU S Y, GAO K, YONG L Q. Improved RT-DETR algorithm for aerial small object detection [J]. Computer Engineering and Applications, 2025, 61(4): 272-281.
- [13] WU T Y, DONG Y K. YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition[J]. Applied Sciences, 2023, 13(24): 12977.
- [14] 席阳丽, 屈丹, 王芳芳, 等. 基于 FEW-YOLOv8 遥感图像目标检测算法[J]. 郑州大学学报(工学版), 2025, 46(4): 62-69.
- XI Y L, QU D, WANG F F, et al. FEW-YOLOv8 based object detection algorithm for remote sensing images[J]. Journal of Zhengzhou University(Engineering Science), 2025, 46(4): 62-69.
- [15] 邵凯, 李浩刚, 梁燕, 等. 跨尺度特征融合的遥感微小目标检测算法[J]. 系统工程与电子技术, 2025, 47(5): 1421-1431.
- SHAO K, LI H G, LIANG Y, et al. Remote sensing tiny target detection algorithm based on cross-scale feature fusion[J]. Systems Engineering and Electronics, 2025, 47(5): 1421-1431.
- [16] 罗显志, 汪航. 跨尺度特征融合的无人机小目标检测算法[J]. 计算机工程与应用, 2025, 61(14): 135-147.
- LUO X Z, WANG H. Small target detection algorithm for UAV based on cross-scale feature fusion [J]. Computer Engineering and Applications, 2025, 61(14): 135-147.
- [17] 张亮亮, 朱海, 卫晓龙, 等. 基于深度学习的航拍影像小目标检测算法[J]. 光电与控制, 2025, 32(11): 47-53, 70.
- ZHANG L L, ZHU H, WEI X L, et al. A deep learning based algorithm for small object detection in aerial images [J]. Electronics Optics & Control, 2025, 32(11): 47-53, 70.
- [18] 刘洋, 时富斌, 王竹筠, 等. 基于显著性引导增强与改进 Faster-RCNN 的遥感图像目标检测方法[J]. 沈阳航空航天大学学报, 2024, 41(6): 50-60.
- LIU Y, SHI F B, WANG Z J, et al. Saliency guided enhancement and improved Faster-RCNN for object detection method in remote sensing images[J]. Journal of Shenyang Aerospace University, 2024, 41(6): 50-60.
- [19] 刘勇, 吕丰顺, 李学琨, 等. 基于 YOLOv8-LGA 的遥感图像目标检测算法[J]. 光电子·激光, 2025, 36(9): 928-937.
- LIU Y, LYU F S, LI X K, et al. Remote sensing image object detection algorithm based on YOLOv8-LGA[J]. Journal of Optoelectronics · Laser, 2025, 36(9): 928-937.
- [20] 吕杰, 杜振丹, 邓佳雯, 等. FFANet:特征融合与注意力机制结合的火烧迹地提取方法[J]. 地球信息科学学报, 2025, 27(12): 2983-2995.

LYU J, DU Z D, DENG J W, et al. FFANet: A method for burned area extraction combining feature fusion and attention mechanism[J]. Journal of Geo-information Science, 2025, 27(12): 2983-2995.

[21] 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: 1-15.

[22] LU X, SUGANUMA M, OKATANI T. SBCFormer: Lightweight network capable of full-size ImageNet classification at 1 FPS on single board computers [C]//IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 1123-1133.

[23] 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.

[24] XIA G S, BAI X, DING J, et al. DOTA: A large-

scale dataset for object detection in aerial images[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.

作者简介

梁钰壘, 硕士研究生, 主要研究方向为深度学习、语音识别处理等。
E-mail: yuman_china@163.com

陈鹏宇, 硕士研究生, 主要研究方向为深度学习、目标检测等。
E-mail: 619847874@qq.com

李菲, 本科, 主要研究方向为目标检测、自然语言处理等。
E-mail: 17267054751@163.com

王烈(通信作者), 教授, 硕士生导师, 主要研究方向为FPGA、图像处理等。
E-mail: lwang@gxu.edu.cn

何广斌, 硕士研究生, 主要研究方向为深度学习、图像处理等。
E-mail: heguangbin99@foxmail.com