

基于 ISC-YOLOv8n 的遥感目标检测算法

刘向举¹ 孙 源^{1*}
LIU Xiangju SUN Yuan

摘 要

针对遥感目标检测中小目标漏检和误检以及目标尺度变化差异明显导致检测精度低的问题,文章提出了基于 YOLOv8n 的遥感目标检测改进算法 ISC-YOLOv8n。首先,结合倒残差移动块与空间和通道协同注意力机制 SCSA,设计了倒残差协同注意力模块 IRSA,提升骨干网络和颈部网络对小目标的关注度;然后,融合可切换空洞卷积 SAConv 与特征提取模块 C2f 设计了 SAC2f 模块,提升网络的多尺度特征提取能力;最后,引入内容感知重组特征模块 CARAFE,优化上采样过程。实验结果表明,改进后的算法在数据集 NWPU VHR-10 和 RSOD 上的 mAP50 分别为 90.6% 和 95.2%,相比基准模型提升了 3.1% 和 2.7%,对比主流算法在精度和检测速度上优势明显,该算法模型在提高检测精度的同时,满足实时性检测需求。

关键词

遥感目标检测;注意力机制;CARAFE;YOLOv8

doi: 10.3969/j.issn.1672-9528.2026.02.003

0 引言

目标检测作为遥感图像分析任务中的关键环节,广泛应用于水体监测^[1]和城市规划^[2]等领域。遥感图像中存在的小尺度目标分布密集且特征信息有限,在检测时易受到背景噪声的干扰进而被漏检和误检。此外,目标尺度变化显著以及成像质量易受天气情况的影响,这些因素增加了检测的难度。解决上述难题对于遥感目标检测的应用意义重大。

遥感目标检测方法分为传统的目标检测算法和基于深度学习的目标检测算法^[3]。其中传统目标检测算法的设计和特征提取步骤极大程度依赖人工干预,这种方式计算复杂度高,对于遥感图像中密集目标的检测准确度不足。目前主流研究倾向于基于深度学习的检测方法,Yuan 等^[4]设计了基于 R-CNN 模型的 StripNet 骨干网络,通过条纹卷积和小核卷积的组合,能够在浅层网络中获得较大的感受野。吕向东等^[5]在 SSD 算法网络中引入 CBAM 注意力机制,用于增强对小目标特征的关注。张宇阳等^[6]在 YOLOv8 算法中引入全局特征金字塔网络 GFPN (global feature pyramid network),用于提高不同尺度特征的融合效率。Zhang 等^[7]提出基于 YOLOv5 的 FFCA-YOLO 算法,设计了特征增强模块 FEM (feature enhanced module),该模块通过采用空洞卷积增强了小目标的局部特征表示,有效抑制背景噪声对小目标的干扰,但仅在单一的数据源进行验证,模型的泛化能力有待提升。

上述方法对提升遥感目标检测的准确性做出了一定的贡献,但仍存在不足之处。为此,提出基于改进 YOLOv8 的遥感目标检测算法 ISC-YOLOv8n。首先,结合倒残差移动块 IRMB^[8] (inverted residual mobile block) 与空间和通道协同注意力 SCSA^[9] (spatial and channel synergistic attention) 设计了倒残差协同注意力模块 IRSA,提升对小目标的关注度;其次,在 C2f 模块的 Bottleneck 结构中引入可切换空洞卷积 SAConv^[10] (switchable atrous convolution),能够自适应调整感受野范围,提升网络对多尺度目标的检测能力。最后,引入轻量级上采样模块 CARAFE^[11],利用核重组的方式优化网络的上采样过程。

1 ISC-YOLOv8n 算法网络结构

ISC-YOLOv8n 算法网络结构如图 1 所示。下面小节将对算法网络中改进的模块进行详细阐述。

1.1 IRSA 模块

倒残差移动块 IRMB 通过卷积核尺度为 1×1 的逐点卷积 PWConv 放缩和扩展特征通道,结合卷积核尺度为 3×3 的深度卷积 DWConv 用于增强对局部特征的提取。该模块利用改进的窗口多头注意力用来捕获全局上下文信息,在每个独立窗口内计算注意力矩阵 (attention matrix),这种计算方式可能会忽略窗口外的小尺度特征信息。

本文在保留倒残差移动块轻量化的卷积组合的基础上,融合空间和通道协同注意力机制 SCSA 构成 IRSA 模块,模块结构如图 2 所示。IRSA 能够利用空间特征信息去引导通道特征的学习,并且加强对局部细节的关注度。

1. 安徽理工大学计算机科学与工程学院 安徽淮南 232001

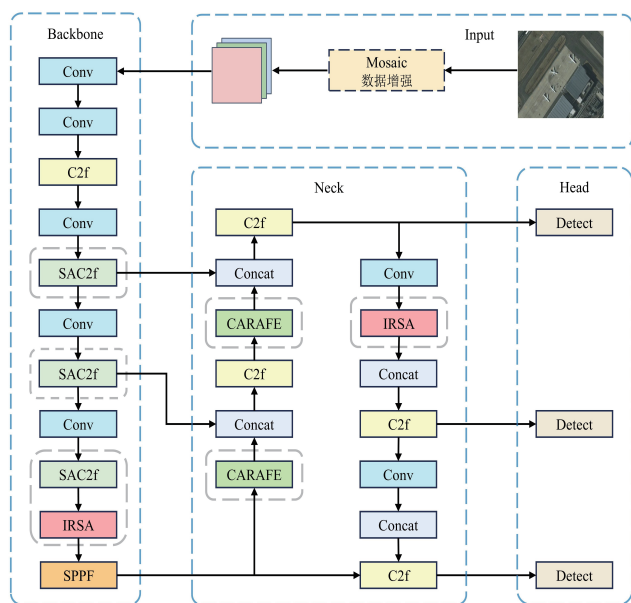


图1 ISC-YOLOv8n 网络结构

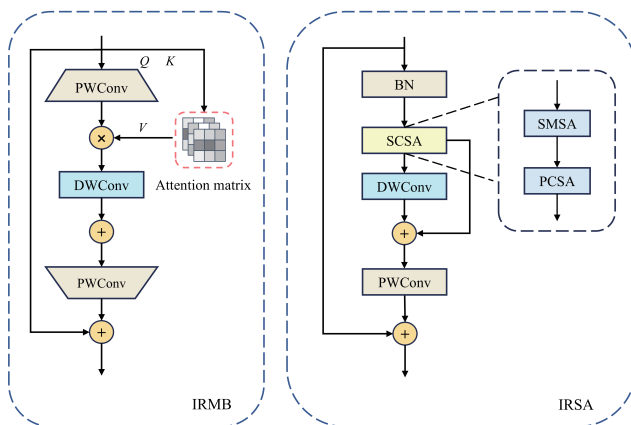


图2 IRSA 模块结构图

SCSA 注意力模块由可共享多语义空间注意力模块 SMSA 和渐进式通道自注意力模块 PCSA 组成, 详细结构如图3所示。

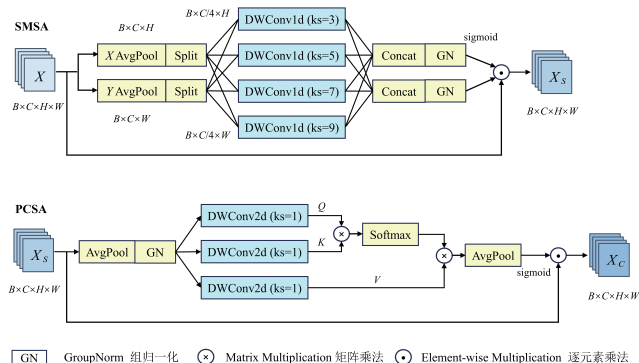


图3 SMSA 和 PCSA 模块结构图

SMSA 模块用于捕捉不同语义层次的空间信息, 首先将输入特征图 $X \in \mathbf{R}^{B \times C \times H \times W}$ 沿 X 轴和 Y 轴平均池化得到单维度特征 $X_H \in \mathbf{R}^{B \times C \times H}$ 和 $X_W \in \mathbf{R}^{B \times C \times W}$, 将单维度特征分解成对应轴

向的子特征 $X_H^i \in \mathbf{R}^{B \times (C/4) \times H}$ 和 $X_W^i \in \mathbf{R}^{B \times (C/4) \times W}$ 。接着对每组子特征应用一维深度共享卷积提取不同尺度的空间信息。最后进行特征拼接, 并通过组归一化和 Sigmoid 激活函数生成空间信息图 $X_S \in \mathbf{R}^{B \times C \times H \times W}$ 。

SMSA 模块生成的空间信息图 X_S 引导后续 PCSA 模块的通道特征学习。PCSA 模块用于优化通道特征并促进跨通道间的特征交互。首先将 SMSA 生成的空间信息图 $X_S \in \mathbf{R}^{B \times C \times H}$ 进行渐进式压缩得到关键空间特征。后使用二维深度卷积将关键特征转换为线性投影, 生成查询 Q (query), 键 K (key), 值 V (value) 特征, 利用这些特征计算通道间的相似性。将计算后的特征通过平均池化和 Sigmoid 激活函数生成优化后的通道注意力图。最后将 SMSA 生成空间注意力图和 PCSA 优化的通道注意力图进行融合得到最终特征表示 X_C 。

1.2 SAC2f 模块

遥感图像中目标尺度差异变化显著, 容易出现误检测进而影响整体目标的检测精度, 为此需要采用能够自适应提取多尺度特征的方法。如图4所示, 可切换空洞卷积 SAConv 能够动态融合多个空洞率卷积来捕获不同尺度的特征信息, 对同一张特征图的不同位置的特征采用自适应的提取策略, 有效处理复杂的特征表示。

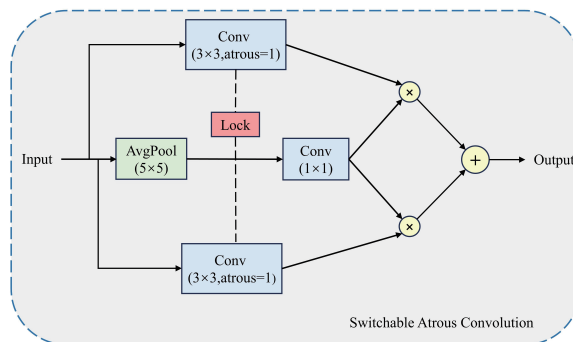


图4 可切换空洞卷积 SAConv 结构图

特征图经过可切换空洞卷积 SAConv 分为两个分支操作, 处理过程如下:

第一条分支将特征图 X 经过卷积核大小为 3×3 的标准卷积操作 $\text{Conv}(x, w, r_0=1)$ 保留原始感受野的大小。通过另一条分支将同一特征图 X 通过卷积核大小为 3×3 的空洞卷积操作 $\text{Conv}(x, w+\Delta w, r_1=3)$ 扩大感受野范围。最后激活开关 (Lock) 融合两个分支的响应信息, 通过平均池化和卷积核大小为 1×1 的卷积操作生成输入和位置相关的自适应权重 $S(x)$, 根据特征图的输入内容动态选择标准卷积和空洞卷积分支的响应强度, 融合公式为:

$$F(x) = S(x) \cdot \text{Conv}(x, w, r_0) + (1 - S(x)) \cdot \text{Conv}(x, w + \Delta w, r_1) \quad (1)$$

式中: w 表示预训练权重; Δw 表示可训练参数; r_0 表示空洞率大小为 1; r_1 表示空洞率, 大小为 3。

本文将 SACConv 与特征提取模块 C2f 中的 Bottleneck 相结合形成 SAC2f 模块，能够提升网络特征提取的灵活性和对多尺度目标的感知能力，SAC2f 模块结构如图 5 所示。

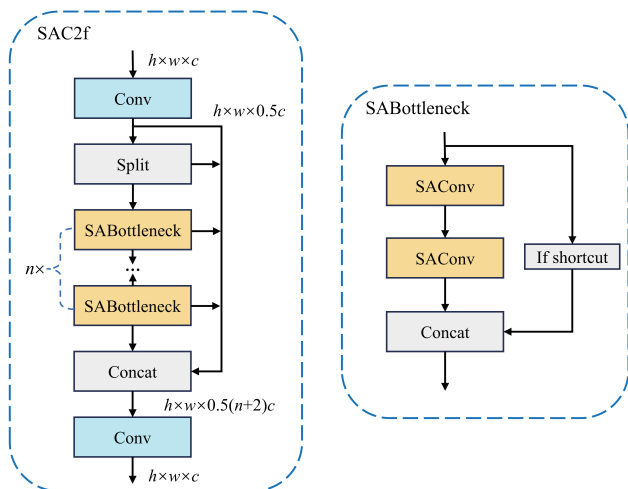


图 5 SAC2f 模块结构图

1.3 CARAFE 上采样模块

本文引入轻量级上采样器 CARAFE，能够根据遥感图像中不同尺度的目标动态生成上采样核。对于小尺度目标，CARAFE 通过内容引导的权重聚焦目标的关键细节，能够有效抑制上采样过程中的细节丢失。CARAFE 包含核预测模块和内容感知重组模块，具体结构如图 6 所示。

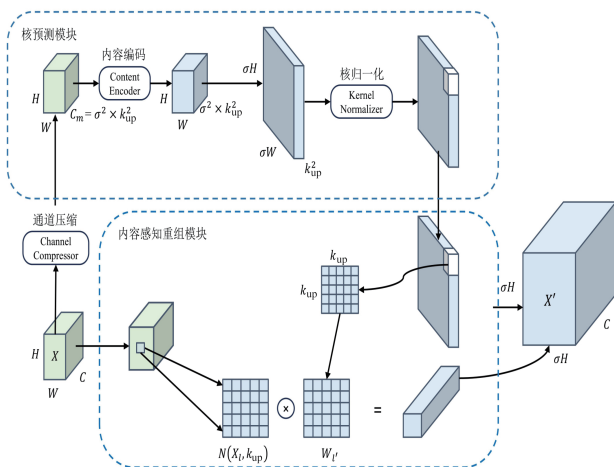


图 6 CARAFE 模块结构图

核预测模块能够根据不同尺度目标的特征重新装配上采样核的大小。首先进行通道压缩 $C \rightarrow C_m$ ，减少后续预测过程的计算量，随后从压缩后的特征图中提取内容用来生成重组核 W_l ，最后使用 softmax 激活函数进行核归一化为后续步骤提供合理的权重分配。

$$W_l = \psi(N(X_l, k_{\text{encoder}})) \quad (2)$$

式中： W_l 表示目标位置 l' 的重组核； $N(X_l, k_{\text{encoder}})$ 表示特征图 X 的局部区域。

内容感知重组模块使用重组核大小为 $k_{\text{up}} \times k_{\text{up}}$ 的特征图进行加权求和，完成特征重组和上采样用于增强特征图的语义信息，使模型能够捕捉到更多的细节信息。

$$X'_l = \phi(N(X_l, k_{\text{up}}), W_l) \quad (3)$$

式中： X'_l 表示特征图目标位置 l' 的像素值； $N(X_l, k_{\text{up}})$ 表示输入特征图 X 的局部区域。

2 数据集准备

实验验证算法的性能以及泛化能力，使用两个开源数据集 NWPUVHR-10^[12] 和 RSOD^[13] 进行实验。数据集均以 8:1:1 的比例被随机划分为独立的训练、验证和测试集。

数据集 NWPU VHR-10 是高分辨率遥感图像数据集，总计 800 幅图像和 3 775 个对象实例，包含飞机、棒球场、篮球场等共计 10 种类别对象。数据集 RSOD 由 936 张注释遥感图像组成，包含飞机、油罐、操场和立交桥 4 类检测目标。以上两种公开数据集包含的目标类型，如图 7 所示。

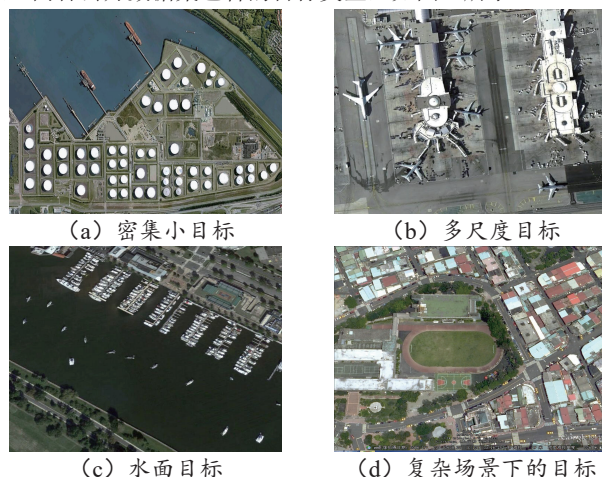


图 7 实验数据集中不同类型目标

3 实验结果与分析

3.1 训练环境与配置

实验环境基于 Ubuntu20.4 系统，使用英伟达 RTX 4090 显卡进行训练。深度学习框架 PyTorch2.0.0，CUDA 版本 11.8，编程语言为 Python3.8。两个数据集采用相同的训练策略，训练超参数设置如下，输入图像尺寸为 640 px×640 px，初始学习率 0.01，动量为 0.937，权重衰减系数 0.000 5，批次大小为 16，训练迭代次数为 200。

3.2 评价指标

平均精度均值 mAP@0.5，参数量 (Params/10⁶) 和检测速度 FPS (Frames Per Second) 来评估改进算法的性能。

3.3 消融实验

消融实验是衡量每一项改进对模型性能影响程度的关

键分析方法。本实验将 YOLOv8n 作为基线算法 (Baseline) 进行训练对比, 其中 “√” 表示在基线模型的基础上引入该模块进行实验。实验确保在相同的训练环境下, 对数据集 NWPU VHR-10 和 RSOD 分别进行消融实验, 实验结果如表 1~2 所示。

表 1 数据集 NWPU VHR-10 消融实验结果表 (B: Baseline)

B	IRSA	SAC2f	CARAFE	mAP@0.5/%	Params/10 ⁶	帧/s
√				87.5	3.01	178.5
√	√			88.7	3.09	168.0
√		√		88.5	3.34	112.1
√			√	88.4	3.15	153.6
√	√	√		89.7	3.42	103.2
√	√	√	√	90.6	3.56	105.3

表 2 数据集 RSOD 消融实验结果表 (B: Baseline)

B	IRSA	SAC2f	CARAFE	mAP@0.5/%	Params/10 ⁶	帧/s
√				92.6	3.01	198.4
√	√			94.1	3.09	160.5
√		√		93.8	3.34	178.3
√			√	94.3	3.15	155.6
√	√	√		94.7	3.42	147.0
√	√	√	√	95.3	3.56	135.7

在数据集 NWPU VHR-10 和 RSOD 的消融实验中, 将 IRSA 模块添加到基线模型的骨干网络和颈部网络中增强了小尺度目标的提取能力, 在两个数据集上 mAP@0.5 分别提高 1.2% 和 1.5%。在添加 IRSA 模块的基础上加入 SAC2f 模块, 在两个数据集上 mAP@0.5 分别提升 2.2% 和 2.1%, 参数量增加了 0.41×10^6 , SAC2f 模块中的可切换空洞卷积 SACConv 可以根据图像中目标特征自适应选择合适的空洞卷积方式, 提升复杂场景下对多尺度目标特征的提取效率, 该计算方式会导致参数量增加。最后加入 CARAFE 上采样模块, 通过内容引导的权重进行特征重组, 在上采样过程聚焦关键细节保留重要特征, 在两个数据集上模型的 mAP@0.5 分别提升 3.1% 和 2.7%, 参数量增加了 0.55×10^6 , 检测速度分别下降 73.2 帧/s 和 62.7 v。

从上述实验结果得出, 改进算法 ISC-YOLOv8n 参数量为 3.56×10^6 相比基线模型增加了 18.2%, mAP@0.5 相较于基线模型在两个数据集上分别提升了 3.1% 和 2.7%, 检测速率分别达到 105.3 帧/s 和 135.7 帧/s, 虽然 FPS 有所下降, 但仍满足实时性检测需求。上述实验性能的表现验证了改进方法的有效性。

3.4 模型对比实验

使用涵盖 10 种不同类别目标的数据集 NWPU VHR-10 在相同的实验环境下, 将本文提出的改进算法 ISC-YOLOv8n

与经典目标检测算法 Faster R-CNN^[14] 以及 YOLOv5、YOLOv8、YOLOv11^[15] 和 YOLOv12^[16] 进行对比试验。此外, 同近年来的遥感目标检测算法 SuperYOLO^[17] 和进行比较。上述算法的对比结果如表 3 所示。

表 3 各模型对比实验结果表

algorithm	mAP@0.5/%	Params/10 ⁶	帧/s
Faster R-CNN	88.7	41.30	21.3
YOLOv5n	85.7	2.51	167.8
YOLOv8n	87.5	3.01	178.5
YOLOv11n	86.8	2.59	201.3
YOLOv12n	83.4	2.57	195.4
SuperYOLO	88.9	4.85	62.4
ISC-YOLOv8n(Ours)	90.6	3.56	105.3

提出的 ISC-YOLOv8n 算法相比双阶段算法 Faster R-CNN, mAP@0.5 提升了 1.9%。同其他经典 YOLO 算法相比, 虽然检测帧数下降, 但在 mAP@0.5 指标上有着显著的优势。与近年来遥感目标检测算法 SuperYOLO 相比, mAP@0.5 提升 1.7%, 在参数量和检测帧数上有着一定优势。通过对比不同算法, 验证了改进算法在参数量没有激增的情况下, 提升检测精度上优势明显。

3.5 实验效果展示

ISC-YOLOv8n 算法在测试集上的检测效果如图 8 所示。在复杂场景下, YOLOv8n 算法将交叉的人行道检测为桥梁, 而 ISC-YOLOv8n 算法避免了这一误检的产生。在密集场景和水面场景的检测表现上, ISC-YOLOv8n 算法改善了目标漏检测的现象, 进一步验证了改进算法的性能优势。

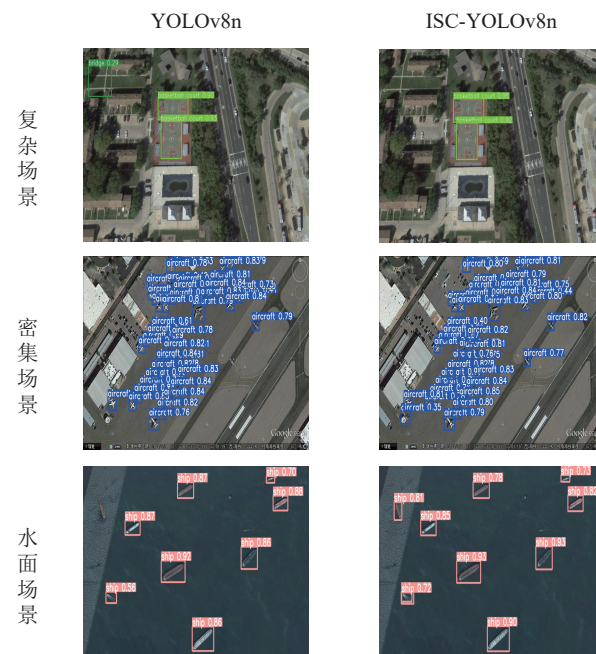


图 8 改进模型的检测效果图

4 总结

本文提出了 ISC-YOLOv8n 算法用于改善遥感图像检测任务中漏检和误检现象,提升了复杂场景下目标识别能力。首先,设计了 IRSA 模块增强对小尺度目标的关注度,用于提升检测精度。其次,结合可切换空洞卷积 SACConv 和 C2f 模块设计了 SAC2f 特征提取模块,提升了多尺度目标的识别效率。最后引入 CARAFE 模块优化网络上采样过程,丰富关键特征。实验结果表明,ISC-YOLOv8n 算法在数据集 NWPU VHR-10 和 RSOD 上取得了较高的精度,平均精度均值达到 90.6% 和 95.2%,相比基准模型分别提升 3.1% 和 2.7%,参数量小幅度增加,且该算法适用于遥感目标检测任务的实时性应用。今后的研究工作将确保模型提升检测精度的同时,优化计算效率进行轻量化改进。

参考文献:

- [1] 刘憬豫,宋开山,刘阁,等.基于中高分辨率遥感影像的水体提取研究进展[J].湿地科学,2025,23(1):72-83.
- [2] 王俊.基于遥感影像助力的城市空间规划系统研究[J].北斗与空间信息应用技术,2025(1):30-33.
- [3] GUI S X, SONG S, QIN R J, et al. Remote sensing object detection in the deep learning era: a review[J]. Remote sensing, 2024, 16(2): 327.
- [4] YUAN X B, ZHENG Z H, LI Y X, et al. Strip R-CNN: large strip convolution for remote sensing object detection[EB/OL].(2025-03-01)[2025-12-30].<https://doi.org/10.48550/arXiv.2501.03775>.
- [5] 吕向东,彭超亮,陈治国,等.基于 RSSD 的遥感图像目标检测算法[J].现代电子技术,2024,47(7):49-53.
- [6] 张宇阳,姜静,符珊.基于改进 YOLOv8 遥感目标检测算法研究[J].通信与信息技术,2025(2):36-40.
- [7] 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.
- [8] ZHANG J N, LI X T, LI J, et al. Rethinking mobile block for efficient attention-based models[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 1389-1400.
- [9] SI Y Z, XU H Y, ZHU X Z, et al. SCSA: exploring the synergistic effects between spatial and channel attention[J]. Neurocomputing, 2025, 634: 129866.
- [10] QIAO S Y, CHEN L C, YUILLE A. DetectorRS: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway:IEEE, 2021: 10213-10224.
- [11] WANG J Q, CHEN K, XU R, et al. CARAFE: content-aware reassembly of features[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway:IEEE, 2019: 3007-3016.
- [12] CHENG G, HAN J W, ZHOU P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS journal of photogrammetry and remote sensing, 2014, 98: 119-132.
- [13] XIAO Z F, LIU Q, TANG G F, et al. Elliptic fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International journal of remote sensing, 2015, 36(2): 618-644.
- [14] 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, 2016, 39(6): 1137-1149.
- [15] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[EB/OL].(2024-10-23)[2025-12-30].<https://doi.org/10.48550/arXiv.2410.17725>.
- [16] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors[EB/OL].(2025-02-18)[2025-12-30].<https://doi.org/10.48550/arXiv.2502.12524>.
- [17] ZHANG J Q, LEI J, XIE W Y, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery[J/OL]. IEEE transactions on geoscience and remote sensing, 2023[2025-12-30].<https://ieeexplore.ieee.org/abstract/document/10075555>.DOI: 10.1109/TGRS.2023.3258666.

【作者简介】

刘向举(1978—),男,黑龙江哈尔滨人,硕士,副教授、硕士生导师,研究方向:物联网、软件定义网络、目标检测。
孙源(1999—),通信作者(email: sy2023vc@163.com),男,安徽蚌埠人,硕士研究生,研究方向:目标检测。

(收稿日期:2025-07-14 修回日期:2026-01-28)