智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

结合倒残差自注意力机制的遥感图像目标检测

赵文清, 赵振寰, 巩佳潇

引用本文:
赵文清, 赵振寰, 巩佳潇. 结合倒残差自注意力机制的遥感图像目标检测[J]. 智能系统学报, 2025, 20(1): 64-72.
ZHAO Wenqing, ZHAO Zhenhuan, GONG Jiaxiao. Remote sensing image object detection based on inverted residual self-attention mechanism[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(1): 64-72.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202312001

DOI: 10.11992/tis.202312001
网络出版地址: https://link.cnki.net/urlid/23.1538.tp.20240923.1520.005

结合倒残差自注意力机制的遥感图像目标检测

赵文清[1,2], 赵振寰[1], 巩佳潇[1]
（1. 华北电力大学 控制与计算机工程学院, 河北 保定 071003; 2. 河北省能源电力知识计算重点实验室, 河北 保定 071003）

摘　要: 针对遥感图像目标检测存在背景信息干扰严重、待检测目标尺寸差异大等问题, 提出一种结合倒残差自注意力机制的目标检测方法。首先, 使用具有强特征提取能力的倒残差自注意力机制骨干网络充分提取目标特征, 降低复杂背景信息的干扰; 其次, 构造多尺度空间金字塔池化模块, 提供多尺度感受野, 增强捕捉不同尺寸目标的能力; 最后, 提出轻量级特征融合模块, 对骨干网络提取的特征图进行融合, 充分结合低层与高层特征, 提高网络对不同尺寸目标的检测能力。与传统网络及其他改进目标检测算法进行对比, 实验发现该方法的检测精度明显优于其他算法。此外, 在 DIOR 数据集和 RSOD 数据集上设计消融实验, 结果表明, 该方法在 DIOR 数据集与 RSOD 数据集上的平均精度均值比 YOLOv8 算法分别提升 4.6 和 4.2 百分点, 明显提升遥感图像目标检测的精度。

关键词: 遥感图像; 目标检测; 倒残差; 自注意力机制; 多尺度; 空间金字塔; 特征提取; 特征融合
中图分类号: TP391　　**文献标志码:** A　　**文章编号:** 1673−4785(2025)01−0064−09

中文引用格式: 赵文清, 赵振寰, 巩佳潇. 结合倒残差自注意力机制的遥感图像目标检测 [J]. 智能系统学报, 2025, 20(1): 64–72.
英文引用格式: ZHAO Wenqing, ZHAO Zhenhuan, GONG Jiaxiao. Remote sensing image object detection based on inverted residual self-attention mechanism[J]. CAAI transactions on intelligent systems, 2025, 20(1): 64–72.

Remote sensing image object detection based on inverted residual self-attention mechanism

ZHAO Wenqing[1,2], ZHAO Zhenhuan[1], GONG Jiaxiao[1]
(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China; 2. Hebei Key Laboratory of Knowledge Computing for Energy & Power, Baoding 071003, China)

Abstract: An inverted residual self-attention method (IRSAM) was proposed in this study as an approach for object detection in remote sensing images. The method was designed to address challenges related to significant variations in object sizes and substantial interference from background information in remote sensing image object detection. Firstly, an inverted residual self-attention mechanism backbone network with strong feature extraction ability was utilized to fully extract the object features, thus reducing the interference of complex background information on the object. Additionally, a multi-scale spatial pyramid pooling module was constructed to offer diverse sensory fields at multiple scales and improve the capacity to detect objects of varying sizes. Finally, a lightweight feature fusion structure was employed to integrate the feature maps extracted from the backbone network, effectively combining low-level and high-level features. The study compared the performance of IRSAM with both traditional network and enhanced object detection algorithms. The results indicated that the proposed method exhibited significantly higher detection accuracy. In addition, ablation experiments were designed on the DIOR and the RSOD datasets. The results show that the mean accuracy is 4.6 and 4.2 percentage points higher than the YOLOv8 algorithm on the DIOR dataset and the RSOD dataset, respectively. Consequently, the proposed method significantly enhances the accuracy of object detection in remote sensing images.
Keywords: remote sensing image; object detection; inverted residual; self-attention mechanism; multi-scale; spatial pyramid; feature extraction; feature fusion

收稿日期: 2023−12−01.　**网络出版日期:** 2024−09−24.
基金项目: 国家自然科学基金项目 (62371188); 河北省自然科学基金项目 (F2021502013); 中央高校基本科研业务费面上项目 (2020MS153, 2021PT018).
通信作者: 赵文清. E-mail: zhaowenqing@ncepu.edu.cn.

遥感图像目标检测旨在定位和分类遥感图像中不同类别的感兴趣目标, 其应用包括: 环境监视、农业监测、地质调查等。由于遥感图像中的

目标存在背景信息干扰严重、待检测目标尺寸差异大等情况, 导致遥感图像目标检测精度较低。虽然基于深度学习的方法已经在该任务中取得了一定的进展, 但是以上问题并未得到很好解决[1]。

现有的基于深度学习的目标检测算法根据是否需要预先生成候选框主要分为两大类, 其中一类目标检测算法称为双阶段目标检测算法, 将生成候选目标框与分类及定位任务划分到两个阶段。此类算法主要以 R-CNN(region-based convolutional neural network)系列算法为代表, 包括 R-CNN[2]、Fast R-CNN[3]、Faster R-CNN[4], 由于这类算法的精度表现较为优秀, 很长一段时间的研究都基于这类算法进行, 在各类任务中都取得了较好的结果。Lu 等[5] 针对图像中的目标旋转角度任意的问题提出了一种基于关键点热力图的双阶段目标检测算法, 增强了对旋转目标的表示能力, 在 DOTA 数据集[6] 上取得了较好的结果。双阶段目标检测算法虽然在精度方面表现较好, 但是由于其过程复杂, 计算量大, 难以应对具有较高速度要求的实时性任务。

单阶段目标检测算法放弃候选目标框生成阶段, 直接对图像进行目标分类及边界框回归, 所以这类算法在速度方面表现较为优秀。单阶段目标检测算法主要以单阶段多框检测算法(single shot multibox detector, SSD)[7] 及 YOLO(you only look once)系列算法为代表。唐嘉潞等[8] 基于单阶段目标检测算法 CenterNet[9] 引入多注意力机制及特征融合, 有效提高小目标检测精度。吴珺等[10] 引入轻量化注意力模块到 YOLO 算法中, 实现轻量化部署。胡硕等[11] 通过使用深度度量学习来改进 YOLOv3[12], 显著改善车辆编号跳变问题。曾文健等[13] 在 YOLOv4[14] 的基础上引入融合非对称特征注意力, 增强网络对图像的特征提取能力。张正等[15] 基于 RetinaNet[16] 提出双向衰减损失方法用于旋转目标检测, 有效提高旋转目标检测精度。曲海成等[17] 提出双向多尺度特征融合方法, 有效提高对遥感图像中不同尺寸目标的检测能力。

此外, 一些学者致力于轻量级小模型的研究, 通过使用一些方法来压缩模型的参数数量、存储空间以及计算复杂度, 从而在保持或提高模型精度的前提下提高目标检测的速度。Sandler 等[18] 提出一种基于深度卷积的与 Resnet[19] 残差结构相反的倒残差模块 (inverted residual block, IRB), 在保证模型准确率的同时使得模型参数量更少, 该模块已成为轻量级模型的基本模块。Mehta 等[20] 结合自注意力机制(self-attention mechanism)与卷积神经网络, 有效融合局部与全局特征, 显著提高了模型的鲁棒性与泛化能力。Zhang 等[21] 提出一种倒移动残差模块, 并构建了一个高效模型 Efficient MOdel, 在保证精度的同时减少了参数量。

以上方法对于常规目标检测具有良好的效果, 但遥感图像不同于普通图像, 将通用目标检测方法应用于遥感图像目标检测领域中尚有不足, 一些针对遥感图像目标检测的方法陆续被提出。郑哲等[22] 提出基于多尺度注意力特征金字塔及滑动顶点回归机制的目标检测算法, 有效解决遥感图像中目标方向随机等问题。赵文清等[23] 先后在 YOLOv5s 的基础上引入 Swin Transformer 及上下文信息加权操作来增强模型的特征融合能力, 有效解决遥感图像中小目标检测困难等问题。提出的基于 YOLOX 算法的多尺度遥感图像目标检测算法[24], 引入自适应空间特征融合与多尺度注意力特征融合模块, 有效提高遥感图像目标检测的精度。Yue 等[25] 提出 SCFNet(semantic correction and focus network), 通过计算图像全局特征与局部特征之间的相似度来对语义特征进行校正, 提高了检测精度。Li 等[26] 提出不同方向、形状及姿态的自适应点表示, 使用自适应点评估和分配样本方案来对点集质量进行衡量, 该模型在遥感图像旋转目标检测任务中获得了较优的性能。Ren 等[27] 基于单层特征提出一种单阶段遥感目标检测方法 StrMCsDet, 有效减少了计算开销。Chen 等[28] 提出了一种一致性和依赖性引导的知识蒸馏方法, 在提高检测精度的同时降低了模型的体积与推理时间。虽然以上方法较好地解决了遥感图像目标检测中存在的目标方向随机、小目标检测困难以及计算开销大等问题, 但是图像背景信息干扰严重、待检测目标尺寸差异大的问题并未得到很好解决。

综上所述, 本文结合遥感图像目标检测中背景信息干扰严重、待检测目标尺寸差异大等问题, 提出一种结合倒残差自注意力机制(inverted residual self-attention mechanism, IRSAM)的目标检测方法。主要工作如下: 首先, 骨干网络使用结合自注意力机制的倒残差骨干(inverted residual self-attention backbone, IRSAB)增强 IRSAM 对目标的特征提取能力。其次, 构造多尺度空间金字塔池化(multi-scale spatial pyramid pooling, MSSPP)模块, 提供不同大小的感受野, 增强所提方法捕捉不同尺寸目标的能力, 并增强特征表达能力。

最后，提出轻量级特征融合模块 OSA-DSCSP（one-shot aggregation distribution shifting cross stage partial），在增加特征多样性及表达能力的同时减少参数量与计算量，提高检测精度。

1　相关技术

1.1　YOLOv8 网络

本文提出的 IRSAM 方法基于 YOLOv8 网络进行改进。YOLOv8 的网络结构主要由特征提取骨干网络、特征融合颈部结构及目标位置与类别预测 3 部分组成。YOLOv8 在将图像输入骨干网络进行特征提取前进行数据预处理，使用 Mosaic 数据增强方法对图片进行随机缩放、裁剪和拼接等操作，增加数据多样性。YOLOv8 在最后 10 epoch 的训练中关闭 Mosaic 数据增强，使模型根据真实数据分布进行微调，提高模型收敛稳定性。

YOLOv8 的骨干网络由修改后的 Darknet-53 网络构成。特征融合部分则采用 PAN（path aggregation network）结构来进行自底向上和自顶向下两个方向的特征融合，从而更好地适应不同尺寸的目标。目标位置与类别预测部分采用解耦头结构来将分类头和定位头分离，使用不同的分支分别进行定位与分类任务，根据边界框预测与分类得分，经过一系列后处理（阈值过滤、尺度还原、非极大值抑制）输出物体的位置与类别信息。

1.2　倒残差结构

倒残差结构采用与残差结构相反的通道升维与降维操作，残差结构先进行通道的降维操作，再通过卷积提取特征，最后进行通道的升维操作。而倒残差结构首先进行通道的升维操作，通过在高维空间使用深度卷积来提取更多特征信息，接下来在对通道进行降维时使用线性激活函数，避免特征信息的丢失，从而提高特征提取能力。

2　网络结构

本文提出的 IRSAM 方法如图 1 所示。在骨干网络中借鉴倒残差移动模块[21]的思想，使用能够提取更多有效信息的倒残差自注意力机制骨干（IRSAB），结合自注意力机制与倒残差模块，增强 IRSAM 对复杂背景信息的过滤能力；构造 MSSPP 模块，提供具备不同尺度感受野的特征图，增强特征表示能力，提高 IRSAM 对不同尺寸目标的检测能力；提出轻量级特征融合模块 OSA-DSCSP 进行特征融合，利用分布移位卷积（distribution shifting convolution, DSConv）[29]减少参数量，同时增强对不同尺寸目标的检测能力，提高所提方法的检测精度。

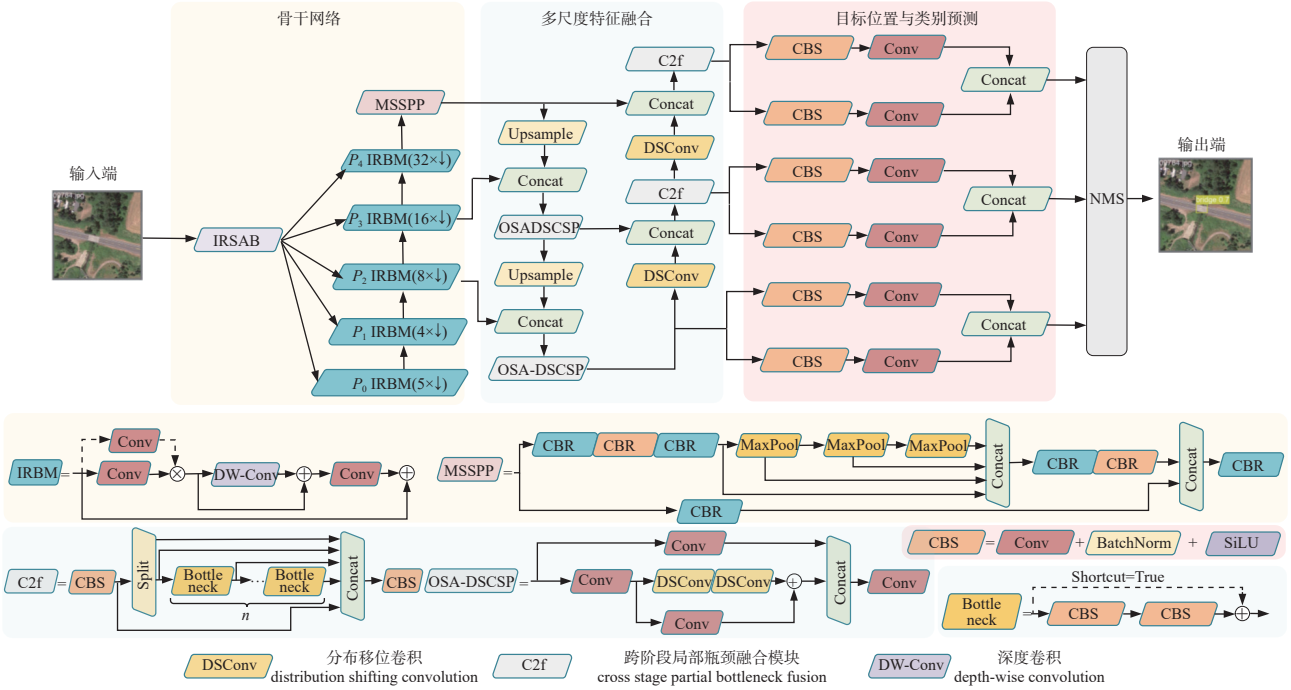


图 1　结合倒残差自注意力机制的目标检测方法结构

Fig. 1　IRSAM target detection method structure

2.1　倒残差自注意力机制骨干网络

遥感图像的背景信息复杂主要是指地物周围存在大量形似物，背景包含大量噪声信息。该问题导致待检测目标的特征与背景特征区分度低，

传统方法进行多次迭代后容易造成待检测目标信息丢失。因此使用结合自注意力机制的倒残差骨干作为骨干网络。首先，IRSAB 结合自注意力机制构建特征之间的语义相关性，通过动态地学习各位置的特征权重，从而突出待检测目标的特征，抑制背景特征的干扰。其次，在 IRSAB 中包含大量跳跃连接分支，直接将输入特征拼接到输出特征上，保证高层输出特征保留低层特征包含的细节信息，减少待检测目标信息丢失。因此，IRSAB 能较好克服遥感图像背景信息复杂以及目标信息容易丢失的问题。

IRSAB 是一个五阶段高效骨干网络，其结构如图 2 所示，IRSAB 输入为遥感图像，经过五阶段的特征提取后输出特征图。相比其他结合自注意力机制的骨干网络，IRSAB 结构简单，仅由倒残差基础模块（IRBM）组成，由于自注意力机制更适合处理网络深层的语义信息[20]，所以仅在第 4 与第 5 阶段开启多头自注意力（multi-head self-attention，MHSA）机制。IRBM 是 IRSAB 中的基础模块，该模块仅包含卷积和自注意力机制，IRBM 的结构如图 3 所示。

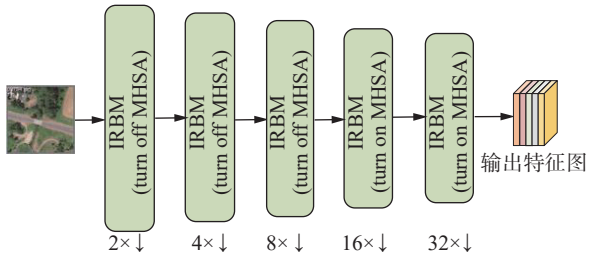


图 2　倒残差自注意力骨干结构

Fig. 2　IRSAB structure

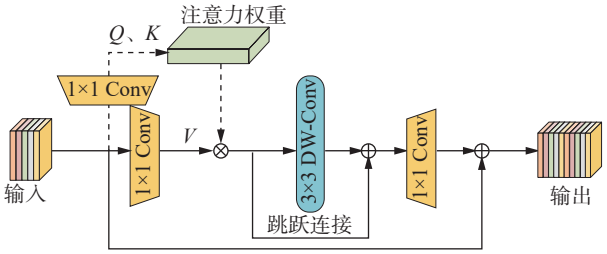


图 3　倒残差基础模块结构

Fig. 3　IRBM structure

倒残差轻量模块既具备卷积的局部建模能力，又拥有自注意力机制的全局建模能力，其结构为级联的卷积操作与自注意力机制，计算过程简化为

$$O = \mathrm{Conv}(\mathrm{MHSA}(X_i)) \tag{1}$$

式中：O表示输出特征图，Conv表示卷积操作，X_i表示输入特征图。

在倒残差轻量模块中，首先对输入的特征图计算其查询、键与值矩阵，使用 $\boldsymbol{Q}$、$\boldsymbol{K}$ 与 $\boldsymbol{V}$ 表示。$\boldsymbol{Q}$、$\boldsymbol{K}$ 与 $\boldsymbol{V}$ 分别通过 1×1 的卷积计算，当需要关断自注意力机制时，关闭计算 $\boldsymbol{Q}$、$\boldsymbol{K}$ 的分支即可。在倒残差结构中（如图 3 中横向分支所示），两个 1×1 卷积分别用于特征图通道升维与降维，在倒残差结构中先对特征图进行通道升维，经过带跳跃连接的 3×3 的深度卷积 (depth-wise convolution, DW-Conv) 后再进行通道降维，最终输出更新后的特征图。深度卷积使用与输入通道数量相同的卷积核并行完成卷积操作，能够有效降低计算复杂度，且在提高模型效率的同时可充分提取各个通道中的信息。

2.2　多尺度空间金字塔池化

本文构造了多尺度空间金字塔池化（MSSPP）模块，该模块可提升所提方法的特征表示能力，如图 4 所示。

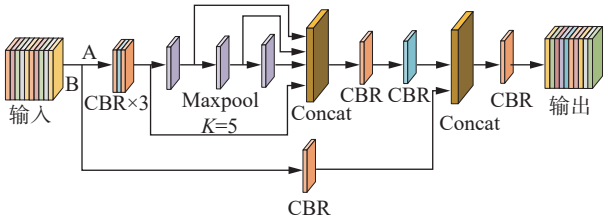


图 4　多尺度空间金字塔池化模块结构

Fig. 4　MSSPP module structure

MSSPP 的输入为 IRSAB 对图像提取得到的特征图，输入特征图在经过 MSSPP 时分别经过两个分支，其中，A 分支用于空间金字塔池化，特征图进入该分支后首先经过 3 个 CBR 模块（卷积、批归一化、ReLU 激活函数），进而输入空间金字塔池化结构，该结构使用 3 个串联的 5×5 最大池化模块形成深度池化网络，以此迭代来增加感受野，捕捉更大范围的特征及上下文信息。拼接经过空间金字塔池化后的 4 个不同尺度特征图，输出拥有多种感受野的特征图。B 分支经过一个 CBR 模块，以此增加反向传播的梯度值，其输出与 A 分支输出拼接，经过一个 CBR 模块输出最终特征图。MSSPP 通过对输入的特征图进行不同尺度的池化，增强特征表示能力，提高所提方法对不同尺寸目标的检测能力，进而有效提高检测精度。

2.3　轻量级特征融合

在骨干网络对遥感图像提取特征的过程中，较低层的特征图包含丰富的细节信息，适用于检测小尺寸目标；随着网络的加深，较高层的特征图包含丰富的语义信息，适用于检测大尺寸目标。为充分融合细节信息与语义信息，本文提出

轻量级特征融合模块 OSA-DSCSP，增加特征多样性及表达能力，增强所提方法的特征融合能力，提高检测精度。

OSA-DSCSP 通过引入分布移位卷积减少参数量，构成跨阶段局部网络提高多尺度聚合能力，其结构如图 5 所示。

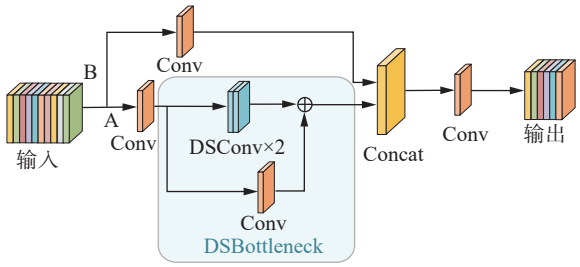


图 5　OSA-DSCSP 结构
Fig. 5　OSA-DSCSP structure

OSA-DSCSP 模块中，输入特征图经过 2 个分支，其中，输入特征图进入 A 分支后，经过一个卷积模块后输入 DSBottleneck 模块。DSBottleneck 使用一次聚合法（one-shot aggregation）聚合不同分支的输出，增加特征多样性。B 分支用于保留原始特征信息。DSBottleneck 的输出与 B 分支的输出按通道维度拼接，经过卷积模块后输出特征图。

如图 1 所示，本文所提方法 IRSAM 使用两个 OSA-DSCSP 模块。第 1 个 OSA-DSCSP 模块的输入为骨干网络 P_3 层输出与经过上采样的 MSSPP 模块输出拼接后的特征图，第 1 个 OSA-DSCSP 模块输出上采样后与骨干网络 P_2 层输出拼接，作为第 2 个 OSA-DSCSP 模块输入。OSA-DSCSP 模块在保留原始特征信息的同时，增加特征的多样性，输出特征图充分融合细节信息与语义信息，有效提高对不同尺寸目标的检测能力。

经过多尺度特征融合层后的特征图将输入检测头，本文所提方法 IRSAM 使用与 YOLOv8 相同的解耦头结构，使用两个独立分支进行目标位置与类别预测。

3　实验结果

3.1　实验环境及参数设置

本实验使用 Windows11 操作系统，GPU 为 NVIDIA GeForce RTX 4070，深度学习框架为 Pytorch2.0.1，CUDA 版本为 11.8。为尽量降低骨干网络初始化权重对实验结果的影响，实验使用的倒残差自注意力机制骨干网络已在 COCO2017 数据集上进行过预训练。实验采用随机梯度下降算法 (stochastic gradient descent, SGD) 进行优化，共训练 300 个 epoch。本实验初始学习率设置为 0.01，最小学习率为 0.001，batchsize 为 16。在训练开始时先进行 3 个 epoch 的 warm-up 训练，动量参数在 warm-up 训练阶段设置为 0.8，warm-up 训练结束后修改为 0.937。

3.2　数据集和评价指标

本实验使用 DIOR 数据集和 RSOD 数据集作为实验数据集。其中，DIOR 数据集包含 23 463 张图像，这些图像被标注了 192 472 个属于 20 个不同目标类别的实例。在 DIOR 数据集中，我们随机选择 11 725 张图像作为训练集，剩余 11 738 张图像作为测试集。RSOD 数据集有 936 张图片，包含操场 (playground)、油桶 (oiltank)、飞机 (aircraft)、立交桥 (overpass)4 个类别，按照 8∶2 的比例进行训练集与测试集的划分，训练集图片数量为 746 张，测试集图片数量为 190 张。

本文使用以下指标来评估性能：精度评估指标采用平均精度（average precision，AP）和平均精度均值（mean average precision，mAP），速度评估指标的单位为 f/s，模型大小评估指标采用参数量（Parameters）。

3.3　实验结果及分析

3.3.1　消融实验结果及分析

本文在 DIOR 数据集与 RSOD 数据集上分别设计 8 组消融实验，证明所采用模块的有效性。在 DIOR 数据集上的实验结果如表 1 所示。

表 1　DIOR 数据集的消融实验结果
Table 1　Ablation results on the DIOR dataset　%

方法	mAP
YOLOv8	77.1
YOLOv8+MSSPP	78.2
YOLOv8+OSA-DSCSP	78.5
YOLOv8+IRSAB	81.0
YOLOv8+MSSPP+OSA-DSCSP	78.6
YOLOv8+IRSAB+MSSPP	81.3
YOLOv8+IRSAB+OSA-DSCSP	81.2
YOLOv8+IRSAB+OSA-DSCSP+MSSPP	**81.7**

表 1 中原 YOLOv8 算法在 DIOR 数据集上 mAP 为 77.1%，本文所提出方法中，加入 MSSPP 模块，mAP 可以提升 1.1 百分点；特征融合修改为 OSA-DSCSP 后，mAP 可以提升 1.4 百分点；骨干网络使用 IRSAB 后，mAP 可以提升 3.9 百分点。所有改进方法应用后，mAP 提升 4.6 百分点。

在 RSOD 数据集上的消融实验结果如表 2 所示。本文所提方法在 RSOD 数据集上的 mAP 为 94.7%，比原 YOLOv8 算法提升了 4.2 百分点。结果表明本文所提方法在 RSOD 数据集上也具有良好的泛化性。

表 2　RSOD 数据集的消融实验结果
Table 2　Ablation results of RSOD dataset　%

方法	mAP
YOLOv8	90.5
YOLOv8+MSSPP	92.4
YOLOv8+OSA-DSCSP	93.1
YOLOv8+IRSAB	93.6
YOLOv8+MSSPP+OSA-DSCSP	93.8
YOLOv8+IRSAB+MSSPP	94.4
YOLOv8+IRSAB+OSA-DSCSP	94.2
YOLOv8+IRSAB+OSA-DSCSP+MSSPP	**94.7**

3.3.2　不同算法在 DIOR 数据集上的对比及分析

为了充分验证 IRSAM 方法的有效性，对本文提出的 IRSAM 方法在 DIOR 数据集上与其他一些主流的目标检测算法进行了实验对比，其结果如表 3 所示。由表 3 可知，相较于两阶段目标检测算法，本文提出的 IRSAM 方法的 mAP 和检测速度有大幅度提升。相较于单阶段目标检测算法 SSD、RetinaNet、CenterNet、YOLOv3、YOLOv4、YOLOv5、YOLOX、YOLOv8 等，mAP 值也有较大提升。与其他改进遥感目标检测方法 AAFNet[24]、SCFNet[25]、StrMCsDet[27] 相比，由于 IRSAM 方法利用 IRSAB 减轻了图像背景信息的干扰，并通过 MSSPP 和 OSA-DSCSP 模块提高了对不同尺寸目标的检测能力，因此其 mAP 值分别提升了 6.4、11.8、16.1 百分点。

表 3　不同算法在 DIOR 数据集上的检测结果比较
Table 3　Comparison of detection results of different algorithms in DIOR dataset

算法	Backbone结构	输入尺寸	mAP/%	检测速度/(f/s)	参数量/10^6
SSD	VGGNet	300×300	55.7	45	26.3
Faster R-CNN	VGGNet	600×1 000	52.1	21	136.7
CenterNet	Resnet50	640×640	55.2	20	32.7
RetinaNet	Resnet50	640×640	61.6	29	37.9
YOLOv3	Darknet53	640×640	58.0	27	61.5
YOLOv4	CSPDarknet53	640×640	65.0	24	52.5
YOLOv5	CSPDarknet	640×640	69.6	42	7.1
YOLOX	CSPDarknet	640×640	72.2	44	8.9
AAFNet[24]	Modified CSPDarknet	640×640	75.3	34	14.1
SCFNet[25]	Modified CSPDarknet	640×640	69.9	14	32.1
StrMCsDet[27]	CSPDarknetC5	608×608	65.6	39	41.4
YOLOv8	CSPDarknet	640×640	77.1	**56**	**3.0**
IRSAM	IRSAB	640×640	**81.7**	50	4.7

此外，由表 3 可知，除 YOLOv8 以外，IRSAM 的参数量比其他通用目标检测算法有较为可观的减少。相较于改进的遥感图像目标检测算法 AAFNet、SCFNet 与 StrMCsDet，IRSAM 的参数量分别减少了 9.4×10^6、27.4×10^6、36.7×10^6。由于 MSSPP 模块引入额外参数量，导致 IRSAM 方法在参数量方面较 YOLOv8 增加了 1.7×10^6，在速度方面较 YOLOv8 降低了 6 f/s，但与其他算法相比，IRSAM 的检测速度仍具有较大优势。

对比本文方法与其他目标检测算法在 DIOR 数据集上每一类目标的检测精度，结果表明该方法对于各类别目标的检测精度均有较大提升，其结果如表 4 所示。其中 c1~c20 表示 DIOR 数据集中 20 个目标类别的检测精度，由表 4 可知，本文方法几乎在所有类别的检测精度都达到了最高的精确度，明显优于其他算法。

表 4　不同算法在 DIOR 数据集的检测结果详细比较
Table 4　Comparison of detailed detection results of different algorithms in DIOR dataset　%

算法	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	mAP
SSD	58.6	67.1	68.1	83.6	26.2	77.2	53.5	67.7	48.2	75.2	56.7	54.3	50.7	34.9	67.9	28.2	77.8	46.0	18.9	52.5	55.7
FasterR-CNN	47.9	64.7	68.6	84.0	23.7	76.4	53.1	57.5	47.0	74.6	56.8	42.0	49.2	16.6	70.5	20.9	73.9	52.9	12.2	49.1	52.1
CenterNet	65.7	64.7	69.2	84.8	25.8	73.8	46.8	54.0	48.0	69.5	56.9	39.7	48.2	45.1	47.6	39.6	79.8	50.3	30.2	65.1	55.2
RetinaNet	71.8	65.7	71.1	87.9	30.9	79.2	57.3	69.9	54.8	79.4	74.1	55.8	53.3	50.1	70.7	40.2	83.8	45.9	21.4	68.4	61.6
YOLOv3	74.7	54.6	69.4	83.8	27.2	73.5	47.7	50.2	46.9	57.7	44.2	57.8	47.3	88.6	29.3	72.3	85.8	27.3	47.4	73.4	58.0

续表 4

算法	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	mAP
YOLOv4	84.8	65.5	74.7	85.1	36.3	78.6	52.3	57.2	54.9	71.3	69.2	58.2	56.2	88.0	38.7	67.8	85.8	49.0	49.9	75.9	65.0
YOLOv5	85.9	76.1	72.3	89.4	43.6	80.8	61.5	59.9	58.0	75.5	73.8	62.1	57.6	89.1	55.7	72.7	86.9	55.5	53.8	82.7	69.6
YOLOX	89.3	72.0	75.3	90.2	47.8	79.3	61.5	60.1	66.2	74.2	76.8	58.1	62.3	89.9	71.1	77.5	89.9	61.0	57.3	83.5	72.2
AAFNet[24]	92.7	81.7	81.0	**90.8**	49.7	81.3	69.9	67.9	70.4	80.4	77.3	64.0	63.2	90.4	68.5	78.0	90.6	65.1	56.6	**85.7**	75.3
StrMCsDet[27]	78.6	58.4	81.3	72.0	38.1	79.2	37.1	49.3	49.5	56.8	62.9	35.5	42.5	54.9	66.0	66.6	80.8	38.3	38.3	34.9	65.6
YOLOv8	93.6	82.4	94.0	82.0	46.0	89.2	61.2	73.9	65.8	76.2	79.7	**72.0**	65.7	93.2	94.6	82.1	92.6	43.7	72.9	80.9	77.1
本文算法	**94.7**	**85.5**	**95.5**	87.1	**52.2**	**91.8**	**75.8**	**80.0**	**70.7**	**83.5**	**82.9**	70.5	**69.7**	**93.9**	**95.3**	**85.4**	**95.0**	**66.3**	**75.9**	82.5	**81.7**

3.3.3　不同算法在 RSOD 数据集上的对比及分析

将本文所提方法与其他主流目标检测算法在 RSOD 数据集上进行实验对比，其结果如表 5 所示。

表 5　不同算法在 RSOD 数据集上的检测结果比较

Table 5　Comparison of detection results of different algorithms in RSOD dataset

算法	Backbone结构	输入尺寸	mAP/%	检测速度/(f/s)	参数量/10^6
SSD	VGGNet	300×300	76.4	46	26.3
Faster R-CNN	VGGNet	600×1 000	80.5	7	136.7
CenterNet	Resnet50	640×640	77.3	19	32.7
RetinaNet	Resnet50	640×640	81.7	22	37.9
YOLOv3	Darknet53	640×640	81.6	30	61.5
YOLOv4	CSPDarknet53	640×640	87.8	28	52.5
YOLOv5	CSPDarknet	640×640	83.6	48	7.1
YOLOX	CSPDarknet	640×640	89.4	49	8.9
YOLOv8	CSPDarknet	640×640	90.5	**68**	**3.0**
IRSAM	IRSAB	640×640	**94.7**	50	4.7

由表 5 可知，与目标检测算法 SSD、Faster R-CNN、CenterNet、RetinaNet 相比，IRSAM 的 mAP 值分别提高了 18.3、14.2、17.4、13.0 百分点，速度分别提高了 4、43、31、28 f/s，参数量分别降低了 21.6×10^6、132.0×10^6、28.0×10^6、33.2×10^6。与 YOLO 系列算法相比，虽然 IRSAM 的速度与参数量不是最优，但 IRSAM 的 mAP 值明显高于其他算法。

3.3.4　可视化结果分析

本文对 DIOR 数据集与 RSOD 数据集上的目标检测结果进行了可视化展示，分别如图 6 与图 7 所示。

在图 6 中，对于具有复杂背景信息的图像，如第 1 张，YOLOv8 受到背景信息的干扰，导致无法精确定位边界不清晰的目标，而 IRSAM 通过 IRSAB 有效抑制了背景信息的干扰，实现了对目标的精确定位。对于第 2 张与第 3 张图像，YOLOv8 对不同尺度特征图的融合能力有限，造成检测时局限于目标的局部特征，发生错检。IRSAM 利用 MSSPP 和 OSA-DSCSP 模块充分融合不同尺度的特征图，正确检测出了目标。从图 7 中可以看出，IRSAM 方法对小尺寸以及较大尺寸目标的检测能力均优于 YOLOv8。

(a) YOLOv8 检测结果

(b) IRSAM 检测结果

图 6　在 DIOR 数据集上的可视化结果

Fig. 6　Visualization detection results on the DIOR dataset

(a) YOLOv8 检测结果

(b) IRSAM 检测结果

图 7　在 RSOD 数据集上的可视化结果

Fig. 7　Visualization detection results on the RSOD dataset

通过对比 YOLOv8 与 IRSAM 的检测结果可以发现, IRSAM 方法有效降低了错检率, 能够有效避免背景信息的干扰, 对复杂背景信息过滤效果较好。此外, IRSAM 方法能够良好适应不同尺寸的目标, 进而显著提升了检测精度。

4　结束语

本文提出了一种结合倒残差自注意力机制的遥感图像目标检测方法, 实现了遥感图像中物体的精确定位与识别, 通过在 DIOR 数据集和 RSOD 数据集上的对比实验分析, 得到如下结论:

1)相较于基线模型, 本文方法在 DIOR 数据集和 RSOD 数据集上的 mAP 值分别提升了 4.6 和 4.2 百分点, 与传统算法和其他改进算法相比, mAP 值也有明显提升。

2)通过对基线模型和本文方法的可视化对比发现, 本文方法较好克服了遥感图像中存在的背景信息干扰严重、待检测目标尺寸差异大等问题。

本文方法仍有改进空间, 虽然检测精度得到了较为可观的提升, 但在检测时速度有轻微下降。在后续的工作中考虑探索效率更高的网络结构, 尽量平衡遥感图像目标检测的精度与速度, 提高实时性。此外, 遥感图像目标检测易受成像条件及环境因素的影响, 单纯利用单模态图像无法很好地解决这一问题, 在后续工作中考虑引入多模态技术针对此问题进行研究。

参考文献:

[1] 石争浩, 仵晨伟, 李成建, 等. 航空遥感图像深度学习目标检测技术研究进展[J]. 中国图象图形学报, 2023, 28(9): 2616−2643.
SHI Zhenghao, WU Chenwei, LI Chengjian, et al. Research progress on deep learning object detection technology for aerial remote sensing images[J]. Journal of image and graphics, 2023, 28(9): 2616−2643.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580−587.

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

[4] REN Shaoqing, HE Kaiming, 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.

[5] LU Dongchen, LI Dongmei, LI Yali, et al. OSKDet: orientation-sensitive keypoint localization for rotated object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1172−1182.

[6] XIA Guisong, BAI Xiang, DING Jian, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3974−3983.

[7] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 21−37.

[8] 唐嘉潞, 杨钟亮, 张淞, 等. 结合显微视觉和注意力机制的毛羽检测方法[J]. 智能系统学报, 2022, 17(6): 1209−1219.
TANG Jialu, YANG Zhongliang, ZHANG Song, et al. Detection of yarn hairiness combining microscopic vision and attention mechanism[J]. CAAI transactions on intelligent systems, 2022, 17(6): 1209−1219.

[9] GUO Chaoxu, FAN Bin, ZHANG Qian, et al. AugFPN: improving multi-scale feature learning for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 12592−12601.

[10] 吴珺, 董佳明, 刘欣, 等. 注意力优化的轻量目标检测网络及应用[J]. 智能系统学报, 2023, 18(3): 506−516.
WU Jun, DONG Jiaming, LIU Xin, et al. Attention-optimized lightweight object detection network and its application[J]. CAAI transactions on intelligent systems, 2023, 18(3): 506−516.

[11] 胡硕, 王洁, 孙妍, 等. 无人机视角下的多车辆跟踪算法研究[J]. 智能系统学报, 2022, 17(4): 798−805.
HU Shuo, WANG Jie, SUN Yan, et al. Research on multi-vehicle tracking algorithm from the perspective of UAV[J]. CAAI transactions on intelligent systems, 2022, 17(4): 798−805.

[12] FARHADI A, REDMON J. YOLOv3: an incremental improvement[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1804−2767.

[13] 曾文健, 朱艳, 沈韬, 等. 面向非对称特征注意力和特征融合的太赫兹图像检测[J]. 中国图象图形学报, 2022, 27(8): 2496−2505.
ZENG Wenjian, ZHU Yan, SHEN Tao, et al. Terahertz image detection combining asymmetric feature attention and feature fusion[J]. Journal of image and graphics,

2022, 27(8): 2496−2505.

[14] BOCHKOVSKIY A, WANG C. Y, LIAO H. M, et al. YOLOv4: optimal speed and accuracy of object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle: IEEE, 2020: 2−7.

[15] 张正, 马渝博, 柳长安, 等. 面向遥感图像旋转目标检测的双向衰减损失方法研究[J]. 电子与信息学报, 2023, 45(10): 3578−3586.
ZHANG Zheng, MA Yubo, LIU Chang'an, et al. Research on bidirectional attenuation loss method for rotating object detection in remote sensing image[J]. Journal of electronics & information technology, 2023, 45(10): 3578−3586.

[16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2999−3007.

[17] 曲海成, 王蒙, 柴蕊. 双向多尺度特征融合的高效遥感图像车辆检测[J]. 计算机工程与应用, 2024, 60(12): 346−356.
QU Haicheng, WANG Meng, CHAI Rui. Efficient remote sensing image vehicle detection with bidirectional multi-scale feature fusion [J]. Computer engineering and applications, 2024, 60(12): 346−356.

[18] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510−4520.

[19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770−778.

[20] MEHTA S, RASTEGARI M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer [EB/OL]. (2021−10−05)[2023−11−26]. https://arxiv.org/abs/2110.02178.

[21] ZHANG Jiangning, LI Xiangtai, LI Jian, et al. Rethinking mobile block for efficient attention-based models[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 1389−1400.

[22] 郑哲, 雷琳, 孙浩, 等. FAGNet: 基于 MAFPN 和 GVR 的遥感图像多尺度目标检测算法[J]. 计算机辅助设计与图形学学报, 2021, 33(6): 883−894.
ZHENG Zhe, LEI Lin, SUN Hao, et al. FAGNet: A multi-scale object detection algorithm for remote sensing images based on MAFPN and GVR[J]. Journal of computer-aided design and computer graphics, 2021, 33(6): 883−894.

[23] 赵文清, 康怿瑾, 赵振兵, 等. 改进 YOLOv5s 的遥感图像目标检测[J]. 智能系统学报, 2023, 18(1): 86−95.
ZHAO Wenqing, KANG Yijin, ZHAO Zhenbing, et al. Remote sensing image object detection based on improved YOLOv5s[J]. CAAI transactions on intelligent systems, 2023, 18(1): 86−95.

[24] ZHAO Wenqing, KANG Yixin, CHEN Hao, et al. Adaptively attentional feature fusion oriented to multiscale object detection in remote sensing images[J]. IEEE transactions on instrumentation and measurement, 2023, 72: 1−11.

[25] YUE Chenke, YAN Junhua, ZHANG Yin, et al. SCFNet: Semantic correction and focus network for remote sensing image object detection[J]. Expert systems with applications, 2023, 224: 119980.

[26] LI Wentong, CHEN Yijie, HU Kaixuan, et al. Oriented reppoints for aerial object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1829−1838.

[27] REN Shougang, FANG Zhiruo, GU Xingjian. A cross stage partial network with strengthen matching detector for remote sensing object detection[J]. Remote sensing, 2023, 15(6): 1574.

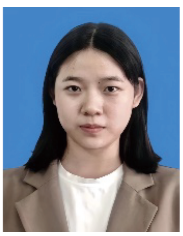
[28] CHEN Yixia, LIN Mingwei, HE Zhu, et al. Consistency- and dependence-guided knowledge distillation for object detection in remote sensing images[J]. Expert systems with applications, 2023, 229: 120519.

[29] NASCIMENTO M G, FAWCETT R, PRISACARIU V A Dsconv: efficient convolution operator[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 5148−5157.

作者简介：

赵文清，教授，博士，主要研究方向为人工智能与图像处理。获河北省科技进步二等奖、三等奖各 1 项。发表学术论文 50 余篇。E-mail：zhaowenqing@ncepu.edu.cn。

赵振寰，硕士研究生，主要研究方向为深度学习与遥感图像处理。E-mail：zhenhuan_zhao@ncepu.edu.cn。



巩佳潇，硕士研究生，主要研究方向为人工智能与图像处理。E-mail：jiaxiao_gong@ncepu.edu.cn。