

文章编号 1004-924X(2025)24-3940-16

融合 Mamba 的多尺度遥感图像实例分割网络

周冬梅, 贾琴琴*

(兰州交通大学电子与信息工程学院, 甘肃 兰州 730070)

摘要: 面对遥感影像中背景场景复杂、目标分布密集且尺度差异导致实例分割精度受限的问题, 提出一种面向全局感知与多尺度特征增强的遥感图像分割算法。设计局部全局感知混频模块, 构建空间及频率双域并行特征通路。空间分支通过多尺度可变形卷积组, 自适应捕获目标边缘细节; 频率分支利用傅里叶变换解耦全局频率成分。该模块可有效增强重叠遮挡复杂场景下的特征判别能力。提出混合卷积自注意力机制, 在保留局部几何特征提取精度的同时, 提升算法对多尺度目标的分割鲁棒性。使用结构增强状态空间模块, 实现空间信息的多维度聚合, 强化复杂背景下小目标的特征表达与分割精度。引入指数移动平均的滑动损失函数, 通过滑动动态加权机制实时调整权重分配, 有效缓解类别不平衡导致的优化偏差问题。在 ISAIID 和 NWPU VHR-10 数据集上的实验结果表明, 所提算法 AP_{50} 分别达到 65.2% 及 68.3%, 较好地兼顾了分割精度和参数量, 综合效果优于诸多先进算法。

关键词: 遥感图像; 实例分割; 状态空间模型; 多尺度特征; 空频域特征融合

中图分类号: TP391.4 **文献标识码:** A

doi: 10.37188/OPE.20253324.3940

CSTR: 32169.14.OPE.20253324.3940

Multi-scale remote sensing image instance segmentation network with mamba integration

ZHOU Dongmei, JIA Qinqin*

(School of Electronic and Information Engineering, Lanzhou Jiaotong University,
Lanzhou 730070, China)

* Corresponding author, E-mail: 1657723517@qq.com

Abstract: Remote-sensing images had low instance-segmentation accuracy due to complex backgrounds, dense targets, and scale gaps. A new segmentation algorithm for global perception and multi-scale enhancement was proposed. A Local-Global Perception Mixer was designed with two parallel pathways. The spatial branch captured target edges via multi-scale deformable convolutions. The frequency branch used Fourier transform to decouple global frequency. These boosted feature discrimination in occluded, complex scenes. Mixed Convolutional Self-Attention was introduced. It retained local geometric-feature accuracy. It also improved multi-scale segmentation robustness. A Structure-Enhanced State-Space module was adopted. It aggregated spatial information multi-dimensionally. It enhanced small-target segmentation accuracy. An Exponential Moving-Average Sliding Loss was proposed. It dynamically adjusted sample weights via a sliding mechanism. This mitigated optimization bias from class imbalance. Experiments

收稿日期: 2025-07-09; 修订日期: 2025-08-08.

基金项目: 国家自然科学基金 (No. 62161016, No. 61661025)

are conducted on ISAIID and NWPU VHR-10 datasets. Its AP_{50} hits 65.2% on ISAIID and 68.3% on NWPU VHR-10. It balances accuracy and model parameters well. It outperforms many state-of-the-art algorithms.

Key words: remote sensing image; instance segmentation; state-space model; multi-scale features; spatial-frequency domain feature fusion

1 引 言

随着遥感成像技术的持续演进,高分辨率遥感影像不断涌现。将这些图像结合人工智能进行高效、精准地理解,对于建筑物提取^[1]、灾害检测^[2]和海洋管理^[3]等各个领域显示出重要的价值。其中光学遥感技术依托高分辨率卫星载荷、无人机多光谱相机等精密光学仪器获取地物反射光谱信息,而地物实例分割精度直接影响其数据的精准解译。实例分割作为计算机视觉中的基本任务之一,凭借从图像中可以提取像素级细粒度目标信息的能力,在处理遥感图像领域得到广泛应用。而与自然图像相比,遥感图像通常从空中向地面拍摄,捕捉到的区域面积大,物体一般呈现出多尺度性差异和背景杂乱等特征,这使传统的实例分割方法在面对遥感图像时难以达到理想的效果。

近年来,研究人员针对遥感图像目标的多样性和特殊性,致力于提升实例分割算法的精度和效率。Chen^[4]在特征聚合步骤中,引入了全局到局部特征交互(Global-to-Local Feature Interaction, GLI)过程,以利用 Transformer 的全局连接上下文,将局部特征连接到 Transformer 全局特征以加强多尺度特征提取。何代毅^[5]将固定步长滑动窗口替换为固定尺寸分割策略,分割所得的每个子窗口均通过多尺度候选框完成感兴趣区域(Region of Interest, ROI)的提取,这种分割方法相当于构建了多种尺度区域生成网络,计算的效率更高。Wei^[6]先构建了低级特征来区分船舶,使用第一级特征提取层补充超分辨率分支并作为低级特征引导网络,得以提升在像素级别捕获区域和纹理信息能力来进行小目标分割。Yasi^[7]通过结合注意力机制与最终预测的相关性来选择性地权衡特征图,使模型专注于前景信息,从而提高检测和分割性能。瑚敏君^[8]等人采用 Mask R-CNN 框架实施算法优化,通过在各级特

征图上额外增加卷积运算,有效提升遥感图像的分辨精度。方俞泽^[9]针对高分辨率遥感影像,优化单阶段实例分割模型 YOLACT,设计出一种结合注意力机制与可变形卷积的轻量化分割算法,显著降低杂乱背景的特征响应。SOLOv2^[10]通过预测实例中心位置来优化小目标分割性能,同时结合多尺度特征融合策略,整合不同层次的特征以增强前景语义表达能力。

Mamba 模型相较于传统 Transformer,在处理高分辨率图像时能优化空间信息建模能力,近期在视觉领域显现优势。VMamba^[11]融合状态空间模型(State-Space Model, SSMs)的优势,提出二维选择性扫描机制(2D Selective Scan, SS2D),整合多方向信息以建立全局感受野,在图像分类、目标检测及语义分割等多类视觉任务中表现出优异的性能和可扩展性。Vision Mamba^[12]应用补丁策略和双向扫描,结合位置嵌入来加强空间信息。LocalMamba^[13]提出了一种搜索高效扫描方向的方法,将输入图像分割成不同的窗口,并分别扫描每个窗口,以增强局部细节提取。

为了实现对遥感图像目标的精准检测和分割,本文对 yolov8n-seg 算法进行改进。改进内容如下:

(1)设计局部-全局感知混频器(Local-global Perception Mixer Block, LGPM Block),构建空间及频率域双域并行特征通路,通过多尺度感受野融合增强复杂形态目标的边缘细节表示,利用傅里叶变换构建全局关系映射,结合位置编码对不同频率成分进行差异化加权,提升模型对目标重叠遮挡等复杂背景的感知能力。

(2)设计混合卷积自注意力(Mixed Convolutional Self-Attention, MCS-Attention),平衡卷积操作的局部特征提取能力与自注意力机制的上下文相关性构建能力,自适应筛选关键特征,进一步强化算法对多尺度地物目标的分割精度。

(3)设计结构增强状态空间模块(Structure-

Enhanced State Space, SESS), 通过构建四条差异化扫描路径, 提升空间信息建模能力并强化关键区域的小目标特征提取。

(4) 引入指数移动平均的滑动损失函数 (Exponential Moving Average Sliding Loss, EMASlideLoss) 该损失函数通过时序窗口动态调整样本权重分布, 并引入动量参数的滑动平均策略, 有效缓解类别不平衡带来的梯度偏差问题。

2 YOLOv8-Seg 算法

YOLOv8-Seg^[14] 作为 YOLOv8 架构的改进型实例分割变体, 其包含: 基于 CSPDarkNet 的主干网络, 集成 C2f 模块与增强型跨层连接, 在保留目标的像素级特征的同时优化梯度传播; 双路径特征金字塔网络 (Feature Pyramid Network, FPN)^[15], 通过双向跨尺度交互实现动态特征融合; 解耦检测头, 将基于 CIOU^[16] 和 DFL^[17] 的无锚框检测分支与原型掩模生成分支相结合, 后者通过系数加权与自适应阈值处理合成实例掩模。该模型在 COCO 基准测试中展现出卓越的精度与效率平衡, 特别适用于图像边缘计算场景。因此本文以 YOLOv8n-seg 通用实例分割模型为基础框架, 针对遥感图像场景特性进行多项算法优化, 以提升目标应用场景下的分割性能。

3 改进算法

3.1 局部-全局感知混频器

如图 1 所示, 输入尺寸为 $H \times W \times 3$ 的遥感图像, 其中 $H \times W$ 表示空间维度。首先经过 3×3 卷积提取低级特征尺寸为 $H \times W \times C$, C 表示通道数量; 随后, 为获取不同尺度的特征表达, 浅层特征连续通过 4 级下采样操作。每一级采用步长为 2 的 3×3 卷积操作, 逐步降低空间分辨率并扩大感受野。经过多层卷积运算后, 原始语义特征会逐步衰减。这一衰减易导致遥感影像中目标的边界模糊与类别混淆。特别当目标尺度差异显著时, 尤为影响实例分割精度。针对该问题本文基于多尺度感受野的双域混合结构^[18], 提出局部-全局感知混频器 (Local-global Perception Mixer Block, LGPM Block), 实现目标形态复杂场景特征的多维度提取与增强。如图 2 所示, 首先, 通过逐点卷积 (Point-wise Convolution, PW-Conv) 将输入特征通道数扩展为两倍, 并沿通道维度均分为尺寸皆为 $H \times W \times C$ 的空间域分支和频域分支, 两个分支分别从局部细节与全局语义层面进行特征提取。

在空间域分支中, 为实现对局部像素关联性与多尺度空间特征的高效提取, 输入特征首先沿通道维度被拆分为两条并行数据流, 以此构建差

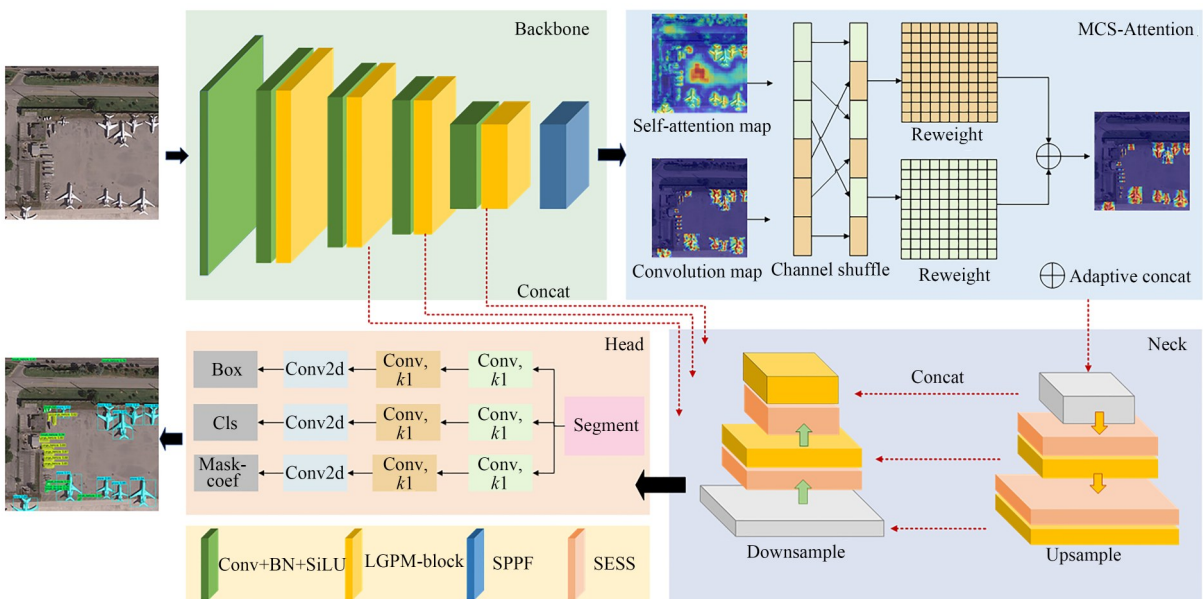


图 1 整体算法结构

Fig. 1 Overall Algorithm Structure

异化特征提取路径。其中一条数据流采用深度可分离卷积 (Depthwise Separable Convolution, DWConv) 进行处理,该操作通过逐通道卷积与点卷积的组合能够精准捕捉局部区域内像素间的细微纹理与几何结构信息;而另一条采用空洞率可配置的深度可分离卷积结构实现区域感受野的自适应扩展。通过并行设置 3 种不同空洞率 (3, 5, 7) 的卷积分支,构建多尺度特征融合模块,避免因传统卷积核采样间隔固定导致的局部特

征丢失问题。最终,将经深度可分离卷积得到的局部纹理特征与提取的多尺度上下文特征进行通道拼接操作,聚合形成包含丰富空间信息的深度特征,其数学表达式为:

$$F_s^d = \text{Concat} [\tilde{f}_{dw}(F_s^1), \tilde{f}_{dw}(F_s^2)], \quad (1)$$

其中: F_s^1 及 F_s^2 为沿通道维度被拆分的两条并行数据流, $\tilde{f}_{dw}(F_s^1)$ 表示对 F_s^1 进行深度卷积得到的局部纹理特征, $\tilde{f}_{dw}(F_s^2)$ 表示对 F_s^2 提取到的多尺度上下文特征。 F_s^d 表示聚合后的深度特征。

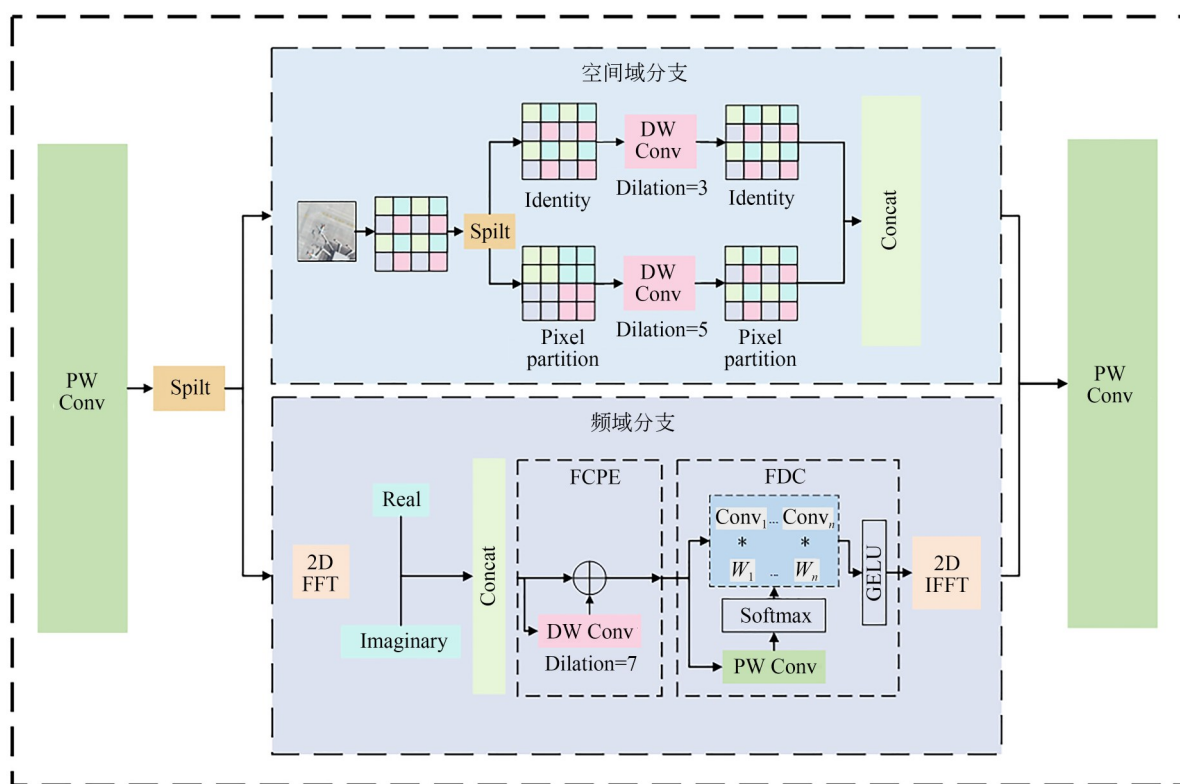


图 2 LGPM Block 模块结构

Fig. 2 LGPM Block module structure

频域分支聚焦于从全局语义层面提取遥感场景的频率特征,有效捕捉目标与背景的结构依赖关系。输入特征首先通过二维快速傅里叶变换 (2D Fast Fourier Transform, 2D FFT) 转换至

$$FFT(a) = \frac{1}{MN} \sum_{a=0}^{M-1} \sum_{b=0}^{N-1} \left[f(a, b) \cdot \exp\left(-i \frac{2\pi}{M} ua\right) \exp\left(-i \frac{2\pi}{N} vb\right) \right], \quad (2)$$

其中: $f(a, b)$ 代表输入图像在空间域坐标 (a, b) 处的像素值; $\exp\left(-i \frac{2\pi}{M} ua\right) \exp\left(-i \frac{2\pi}{N} vb\right)$ 为 FFT 的复指数项,用于在频域中计算特定频率分

频域,从而有效提取图像的周期性结构规律、高频细节特征与全局结构信息。令输入特征图为 a , 当频域坐标取 (u, v) 时,该特征图经 FFT 转换后的频域输出如下:

量; i 为虚数单位; M 和 N 分别对应输入图像的高度与宽度,即 $f(a, b)$ 的行数和列数。

为实现对不同频率分量的精准学习,引入频率条件位置编码 (Frequency-Conditional Position-

al Encoding, FCPE)机制^[18]。FCPE 对输入特征执行幅度提取操作,为网络提供图像的纹理信息与周期性结构,同时降低噪声及冗余信息对特征提取过程的干扰。频域特征的幅度值可反映特定频率分量的强度,其具体计算公式如下:

$$DW(a) = \sqrt{((Re(FFT(a)))^2 + (Im(FFT(a)))^2)}, \quad (3)$$

其中: $FFT(a)$ 表示对输入 a 进行 FFT 变换后的频域输出; $Re(FFT(a))$ 和 $Im(FFT(a))$ 分别为 $FFT(a)$ 的实部和虚部。对提取的幅度值引入 ReLU 激活函数,消除 FFT 过程中产生的负幅度值,仅保留有效频域成分,从而增强网络训练的稳定性。

将幅度与相位融合,网络可更精准地重构图像中的关键结构、纹理及细节,二者共同形成图像的特征信息,进而提升网络的特征表示能力:

$$FCPE(a) = DW(a) \exp \left(i \arctan \left(\frac{Im(FFT(a))}{Re(FFT(a))} \right) \right), \quad (4)$$

其中: $\arctan \left(\frac{Im(FFT(a))}{Re(FFT(a))} \right)$ 代表频域数据的相位信息, $DW(a)$ 代表频域数据的幅度; $\exp \left(i \arctan \left(\frac{Im(FFT(a))}{Re(FFT(a))} \right) \right)$ 为复指数形式的

$$IFFT(a) = \frac{1}{MN} \sum_{a=0}^{M-1} \sum_{b=0}^{N-1} \left[FCPE(a) \cdot \exp \left(i \frac{2\pi}{M} ua \right) \exp \left(i \frac{2\pi}{N} vb \right) \right], \quad (9)$$

其中: $\exp \left(i \frac{2\pi}{M} ua \right)$, $\exp \left(i \frac{2\pi}{N} vb \right)$ 对应 $IFFT$ 的复指数项,用以将频域数据转换至空间域。

因此 LGPM Block 可有效整合频域信息,增强网络对图像中细微特征与频域信息的捕捉能力,提升特征细节的表达效果;并利用 FFT 的去噪作用减少噪声对特征提取的干扰,强化图像特征感知能力,进而提高网络的特征提取水平。

3.2 混合卷积自注意力

在高分辨率遥感图像实例分割任务中,目标数量众多且分布密集,加之尺度多样,对网络的上下文感知能力提出较高要求。传统卷积操作通过滑动窗口对输入特征进行局部感知,具有在捕捉目标几何细节方面优势,但难以有效捕捉上下文相关性信息;而自注意力机制凭借其上下文

相位信息,其功能是保持频域数据的相位状态。

在此基础上,采用频率动态卷积(Frequency-Conditional Positional Encoding, FDC)^[18]进一步挖掘频率特征的多尺度特性,可表示为:

$$b_i = PWConv_i(a) = PWConv2d(a, k_i), \quad (5)$$

其中: i 代表各尺度的卷积操作及对应处理步骤, $k_i \in \{3, 5, 7\}$ 为卷积核尺寸, $PWConv2d$ 指二维卷积操作, $PWConv_i$ 为采用不同大小卷积核后生成的特征图;多个卷积核所生成的多尺度特征集合为 $\{b_3, b_5, b_7\}$ 。通过全连接层(Fully Connected Layer, FC)创建生成每个尺度的通道描述符:

$$W = FC(b_i). \quad (6)$$

计算所有全连接层变换,归一化(Softmax)后得到第 k 个卷积核的权重:

$$wt(b_i)_k = Softmax(FC_k(Z)). \quad (7)$$

最后一步基于权重实现特征的动态选择,通过全连接层生成的注意力权重 W 指引网络关注图像内最相关的尺度特征,随后进行加权求和以减少信息冗余,得到最终的加权特征输出:

$$V = \sum_k (Softmax(FC_k(Z)) b_i). \quad (8)$$

处理后的频率特征经二维逆傅里叶变换(2D Inverse Fast Fourier Transform, 2D IFFT)转换回空间域,结合空间域分支输出:

表征能力,能够动态计算输入特征任意位置间的空间依赖权重,有效学习跨区域语义关联^[19],但模型参数量大。为融合二者优越性,本文设计出高效的混合卷积自注意力(Mixed Convolutional Self-Attention, MCS-Attention),用于提取更具判别力的遥感图像特征。如图 3 所示,在输入特征图与输出特征图中,对于像素坐标 (x, y) , $f_{xy} \in R^{C_n}$ 与 $g_{xy} \in R^{C_n}$ 分别为输入、输出特征图对应位置的特征向量。在 MCS-Attention 的自注意力分支中,首先通过 1×1 卷积操作生成查询(Query)、键(Key)和值(Value)三个投影矩阵,具体计算如式(10)所示:

$$g_{xy} = \sum_{a, b \in N_l(x, y)} A[W_q^{(l)} f_{xy}, W_k^{(l)} f_{ab}] W_v^{(l)} f_{ab}, \quad (10)$$

其中: $W_q^{(l)}$, $W_k^{(l)}$, $W_v^{(l)}$ 分别为第 l 层卷积层计算得

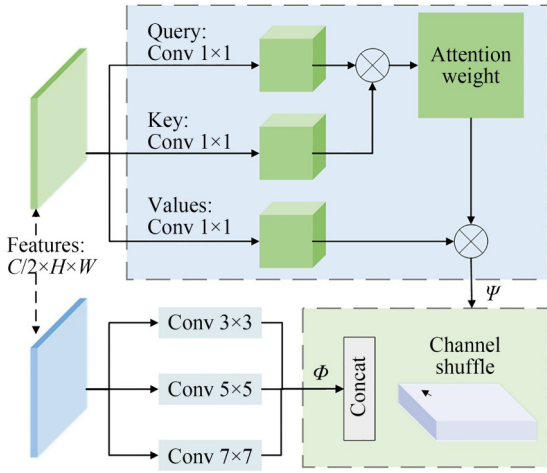


图 3 MCS-Attention 模块结构

Fig. 3 MCS-Attention module structure

到的查询、键和值的权重矩阵; $N_s(x, y)$ 定义以 (x, y) 为中心、大小为 $s \times s$ 的局部空间邻域,用于限定注意力计算范围,符合遥感图像中同类地物在空间分布上的局部聚集特性; $\mathbf{W}_q^{(l)} f_{xy}$, $\mathbf{W}_k^{(l)} f_{ab}$ 对应 $N_s(x, y)$ 位置特征注意力权重 $A(q, s)$ 。公式如下:

$$A(q, s) = \text{softmax} \left[\frac{(\mathbf{W}_q^{(l)} f_{xy})^T (\mathbf{W}_k^{(l)} f_{ab})}{\sqrt{d}} \right], \quad (11)$$

其中: d 为 $\mathbf{W}_q^{(l)} f_{xy}$ 的特征维数,通过引入 $\sqrt{d}$ 进行缩放,可有效避免 softmax 函数梯度消失问题。将自注意力分支输出记为 F_{att} 。

为平衡卷积操作的局部特征提取能力与自注意力机制的上下文相关性构建能力, MCS-Attention 模块最终输出 F_{out} 由卷积路径输出 F_{conv} 与自注意力路径输出 F_{att} 加权融合而成:

$$F_{\text{out}} = \phi F_{\text{conv}} + \psi F_{\text{att}}, \quad (12)$$

其中: ϕ 和 ψ 为两个输出路径分配不同的自适应权重,以适应遥感图像实例分割中不同深度特征的表达需求。此外,在模块输出端引入通道混洗 (Channel Shuffle)^[20], 该操作通过打乱通道顺序, 增强不同通道间的信息交互, 能够有效抑制遥感图像中复杂背景噪声干扰, 同时自适应筛选关键特征通道, 进一步提升网络对多尺度地物目标的分割精度。

3.3 结构增强状态空间模块

针对现有颈部特征融合机制特征感知范围受限以及对小目标和遮挡目标分割精度不足等

挑战^[21], 本文提出一种基于状态空间模型的高效特征融合机制, 深度融合 Mamba 模型的隐藏状态表示, 显著提升特征融合的效率与小目标分割任务的准确性。

该结构增强状态空间模块 (Structure-Enhanced State Space Module, SESS) 基于二维选择性扫描 SS2D 技术, 设计对角线蛇形扫描 SESS 模块, 如图 4, 而模块核心的多方向扫描路径设计细节见图 5。

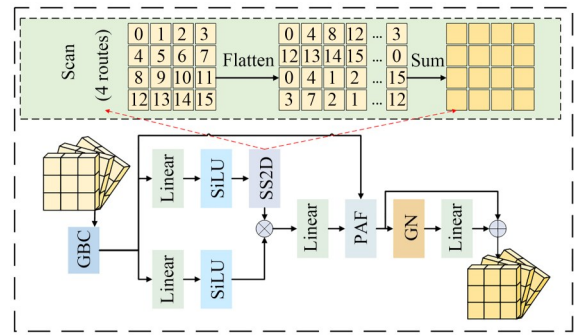


图 4 SESS 模块结构

Fig. 4 SESS structure

针对传统双向扫描和蛇形扫描在捕捉对角线纹理与复杂交织结构时存在的信息缺失问题^[22], 图 5 所示 SESS 通过构建四条差异化扫描路径——两条平行蛇形路径与两条对角蛇形路径, 实现对图像纹理特征的多方向覆盖与深度挖掘。这种设计不仅能够有效提取规则目标区域的连续语义信息, 还能在多个方向上保持纹理的空间一致性, 适用于处理背景复杂、小目标数量众多的遥感图像实例分割任务。

图 4 所示为进一步优化特征表达, 模块前端嵌入低秩近似瓶颈卷积模块 (Generalized Bottleneck Convolution, GBC)^[23], 通过参数压缩与特征重组, 对 SESS 各层间的输出进行精细化处理。GBC 利用低秩矩阵分解技术降低计算复杂度, 同时增强实例级语义特征的提取能力, 显著提升模型对小尺寸目标和被遮挡对象的分割性能。在数据处理流程中, 遥感图像经 Patch Embedding 和 Position Encoding 转换为特征序列后, 输入至 SESS 模块。其过程基于以下公式:

$$\overline{P} = e^{\Delta P}, \quad (13)$$

$$\overline{Q} = (\Delta P)^{-1} (e^{\Delta P} - I) \cdot \Delta Q, \quad (14)$$

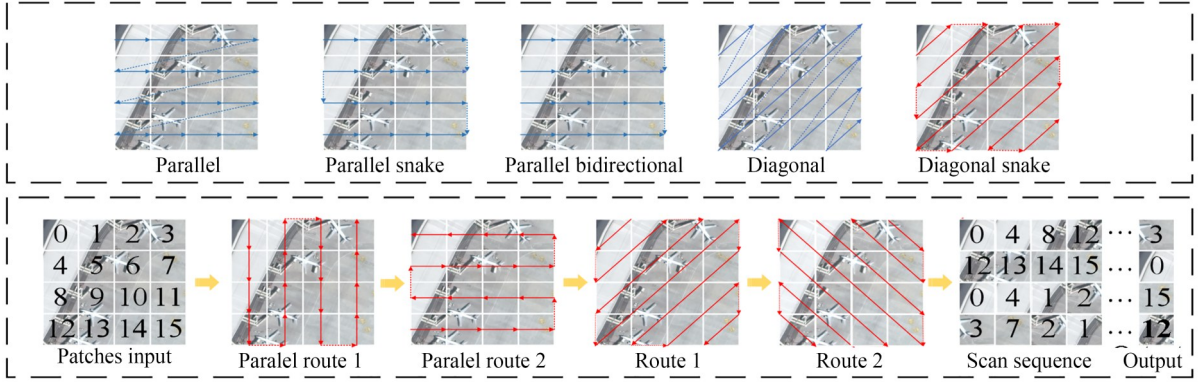


图5 单一扫描路径及 SESS 路径

Fig. 5 single scanning path and SESS path

$$z_k = \overline{P}_{z_{k-1}} + \overline{Q}_{w_k}, \quad (15)$$

$$u_k = R_{z_k} + S_{w_k}, \quad (16)$$

其中:输入序列 $w \in R^{t \times D}$, 矩阵 $P \in R^{G \times D}$ 用于控制隐藏状态的空间分布, $S \in R^{D \times D}$ 构建输入的跳跃连接以保留原始特征信息, z_k 表示时间步长 k 的具体隐藏状态, 通过选择性扫描 SS2D 生成的矩阵 $Q \in R^{G \times D}$ 和 $R \in R^{G \times D}$ 分别是编码空间与时间维度的特征信息。输出 u_k 整合了多方向扫描获取的特征, 通过 SASS 模块建立的多方向邻接关系, 实现对复杂拓扑结构和纹理细节的高效捕捉。

其中为强化模型对遥感图像中目标形状与纹理细节的感知能力, 引入像素注意力导向融合 (Pixel Attention-oriented Fusion, PAF) 机制^[24]。该机制通过动态计算像素级注意力权重, 引导 SESS 模块聚焦于关键区域特征的融合。在选择性扫描完成后, 结合残差连接策略, 确保特征信息在融合过程中的完整性与流动性, 有效避免信息丢失与梯度消失问题, 从而显著提升遥感图像小目标实例分割的精度与鲁棒性。

3.4 指数移动平均的滑动损失函数

交叉熵损失函数是分类任务中广泛采用的一种损失函数, 它能够有效地衡量模型预测的概率分布与真实标签之间的差异。YOLOv8 采用二元交叉熵损失 (Binary Cross-Entropy Loss, BCELoss) 作为分类损失函数, 如式 (17) 所示。然而, BCELoss 对类别不平衡问题非常敏感^[25]。当训练数据中存在显著的正负样本比例失衡, 即前景目标数量远少于背景区域时, 优化过程会因负样本主导梯度更新而产生优化偏差, 导致损失

函数倾向于优先最小化占比较大的负样本误差, 同时抑制正样本的特征学习, 这种现象在复杂场景下的遥感图像中尤为显著。

$$L_{\text{BCE}} = -y \log(\hat{p}) - (1 - y) \log(1 - \hat{p}), \quad (17)$$

其中: L_{BCE} 表示二进制交叉熵损失, 其中 $\hat{p}$ 表示每个样本的预测概率, y 表示样本的真实标签。

$$L_{\text{ESL}} = L_{\text{BCE}} \times ES(x, \mu). \quad (18)$$

本文采用指数移动平均的滑动损失函数 (Exponential Moving Average Sliding Loss, EMASlideLoss)^[25] 作为损失函数。 L_{ESL} 为 EMASlideLoss, $ES(x, \mu)$ 为样品损失权重系数。其中滑动加权函数和指数移动平均函数分别如式 (19) 和式 (20) 所示:

$$ES(x, \mu) = \begin{cases} 1, & x \leq \mu - 0.1 \\ e^{1-\mu}, & \mu - 0.1 < x < \mu + 0.1 \\ e^{1-x}, & x \geq \mu + 0.1 \end{cases}, \quad (19)$$

其中: 预测框与真实框的交并比 (IoU) 记为 x ; 权重阈值 μ 由指数移动平均函数动态更新, 用于区分样本难度。当 $x \leq \mu - 0.1$ 时, 视为简单样本; 当 $x \geq \mu + 0.1$ 时, 视为困难样本; IoU 介于两者之间的边界样本则通过滑动加权函数进行加权。该策略促使模型更聚焦于难以分类的样本, 有效提升了对少数类的识别能力。

$$\begin{cases} \mu_n = \beta \mu_{n-1} + (1 - \beta) \mu_n \\ \beta = \beta_0 (1 - e^{-\frac{n}{\tau}}) \end{cases}, \quad (20)$$

其中: n 代表损失计算更新次数; μ_n 为当前所有样本的平均 IoU; μ_{n-1} 表示前一次更新的平均 IoU; β_0 是初始指数移动权重衰减率; β 为指数移动平

均衰减因子; τ 是控制时间衰减速率的时间常数。

该损失函数通过滑动动态加权机制根据样本 IoU 实时调整权重分配,并结合指数移动平均机制,利用衰减因子 β 平滑历史梯度影响。这种设计不仅通过 EMA 机制有效缓解训练震荡,还利用滑动权重自动平衡正负样本贡献,从而动态调整损失曲面,进一步提高训练稳定性和模型性能。

4 实验结果分析

4.1 数据准备及实验环境配置

ISAID^[26]是遥感领域的大规模数据集,由 JL-1 和 GF-2 卫星拍摄的图像构建而成。该数据集包含 2 806 张原始高分辨率遥感图像,经裁剪为 800×800 大小并切片后,得到 7 500 张训练集图像、1 653 张验证集图像和 1 315 张测试集图像;共有 15 个目标类别。

NWPU VHR-10^[27]数据集是由西北工业大学于 2014 年创建。该数据集包含从 Google Earth 和 Vaihingen 提取的 800 幅高分辨率彩色和全色红外图像。

本文模型训练环境配置为:Ubuntu22.04, GPU 为 NVIDIA GeForce RTX 4090 (24 GB), CPU 为 16 vCPU Intel (R) Xeon (R) Platinum 8481C, Python 版本为 3.10, Pytorch 版本为 2.1.2, CUDA 版本为 11.8。实验均训练 300 个 epoch,批量大小设置 32,图像尺寸设置 640,初始学习率设置为 0.01,动量为 0.937,使用权重衰减率为 0.000 5 的随机梯度下降优化器 SGD。

4.2 评价指标

为了验证本文提出的模型应用于航拍图像检测的有效性,采用准确率(Precision, P),召回率(Recall, R),浮点运算数(Giga Floating-point Operations Per Second, GFLOPs),所有目标类别的平均精度均值(mean Average Precision, mAP),等指标来衡量模型性能。准确率、召回率、mAP 的公式如下:

$$Precision = \frac{TP}{(FP + TP)}, \quad (21)$$

$$Recall = \frac{TP}{(FN + TP)}, \quad (22)$$

$$AP = \int_0^1 Precision(Recall) dRecall, \quad (23)$$

其中:TP 代表正确检测到的目标个数;FP 代表检测到目标但非正确目标的个数;FN 代表目标漏检的个数。

4.3 实验结果与分析

4.3.1 LGPM Block 模块有效性分析

为研究 LGPM Block 空间域分支和频域分支各自对模型性能的贡献,在 ISAID 数据集以 YOLOv8n-seg 为基线模型分别评估以上分支,结果如表 1 所示。

表 1 空频双分支消融实验

Tab. 1 Spatial-frequency dual-branch ablation experiment

Model	Params/M	AP_{50}	AP_s	AP_m	AP_l	FPS
基线	3.26	61.4	42.8	45.9	50.5	104
+S	3.62	62.5	43.5	47.1	51.2	97
+F	3.58	62.1	42.6	46.8	51.8	105
+S+F	3.93	63.6	44.2	48.5	52.3	112

由表中数据可知,在 S 表示的仅空间分支的 LGPM Block,参数量较基线模型增加 0.36 M, AP_{50} , AP_s , AP_m 分别提高 1.1%, 0.7%, 1.2%。在图 6 中说明仅空间分支能有效捕捉遥感图像中小目标的局部边缘细节,但 AP_l 仅提高 0.7%,表明其全局语义感知不足。在 F 表示的仅频域分支的 LGPM Block,参数量降至 3.58M, AP_l 提高 1.3%,能通过傅里叶变换构建全局频率关联以强化大目标特征,但 AP_s 减少 0.2%,图 6 中显示小目标边缘模糊、漏检率升高,其局部细节提取能力存在局限。完整双分支 LGPM Block 较基线模型 AP_{50} , AP_s , AP_m , AP_l 分别提高 2.2%, 1.4%, 2.6%, 1.8%,参数量仅增 0.67 M,推理速度提升 7%,充分证明该模块为遥感图像多尺度目标特征提供高效分割支撑。

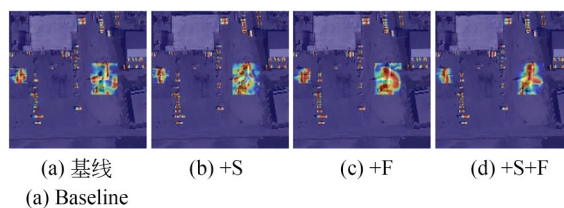


图 6 空频双分支消融实验热力图

Fig. 6 Spatial-frequency dual-branch ablation experiment heatmap

此外,为探究 LGPM Block 模块在遥感图像实例分割任务中的有效性,在 ISAID 数据集展开一组双域优化模块的对比实验。如表 2 所示分别嵌入多尺度双表示对齐滤波器 MDAF^[28]、场景耦合注意力 SCA^[29]、稀疏感知光谱增强模块 SSEM^[30]和本文的局部-全局感知混频器 LGPM Block。

从实验数据可以看出,不同双域特征增强模块在遥感图像多尺度目标分割性能上存在显著差异。其中, LGPM Block 在 AP_{50} , AP_s , AP_m , AP_l 及 FPS 指标上均表现优异。相比于基线模型, LGPM Block 在 AP_{50} 上提升了 2.2%, AP_s 提升了 1.4%, AP_m 提升了 2.6%, AP_l 提升了 1.8%,有效提高了模型的全尺度分割精度。此外, LGPM Block 在参数量方面仅较基线增加 0.67 M,保持了较低的资源消耗水平;而推理速度提升 8 FPS(帧/秒),达到 112 FPS。这表明 LGPM Block 在保持资源消耗可控的同时,成功提升了模型的分割性能。对比其他主流双域模块, LGPM Block 表现出更高的全尺度精度和更优的推理效率。这些结果表明了 LGPM Block 的有效性,它在提升分割精度的同时优化了推理速度,为遥感图像多尺度目标分割任务提供了一个高效的选择。

表 2 各双域模块对比实验结果

Tab. 2 Comparative experimental results of various dual-domain modules

Model	AP_{50}	AP_s	AP_m	AP_l	FPS
基线	61.4	42.8	45.9	50.5	104
+ MDAF ^[28]	61.1	42.6	46.2	50.2	97
+ SCA ^[29]	61.0	43.1	45.6	49.8	101
+ SSEM ^[30]	61.2	42.5	45.7	50.7	100
+ LGPM Block	63.6	44.2	48.5	52.3	112

4.3.2 不同注意力机制的对比实验

为验证 MCS-Attention 模块的实际效能,在 ISAID 数据集以基线模型基础上开展对比实验,分别嵌入高效多尺度注意力 EMA^[31]、自适应细粒度通道注意力 AFG^[32]、全局注意力机制 GAM^[33]和通道-空间注意力 CBAM^[34]。其中,如表 3 数据所示, MCS-Attention 模块展现出轻量化、全尺度精度最优、性能提升显著的核心优势:

与其他注意力机制相比,其参数量与基线模型完全一致,均为 3.26 M,避免了 EMA, GAM 等模块因参数量激增导致的效率损耗;同时,其 AP_{50} 指标达 62.7%,较基线提升 1.3%,不仅显著高于 EMA, AFG, CBAM,也优于仅能维持性能稳定的 GAM。在细分尺度指标上, MCS-Attention 的 AP_s , AP_m , AP_l 均为所有对比模块中的最高值,尤其在中目标和大目标分割上优势突出,充分验证了其在特征优化中的有效性。

表 3 不同注意力机制对比实验

Tab. 3 Comparison of different attention mechanisms

Model	Params/M	AP_{50}	AP_s	AP_m	AP_l
基线	3.26	61.4	42.8	45.9	50.5
+ EMA ^[31]	12.1	61.2	40.4	47.0	50.1
+ AFG ^[32]	3.78	62.1	43.3	44.4	49.7
+ GAM ^[33]	4.62	62.4	42.6	47.8	50.0
+ CBAM ^[34]	3.26	61.7	42.2	45.1	51.7
+ MCS-Attention	3.26	62.7	43.6	48.2	52.1

图 7 的热力图结合表 3 数据显示,相较于基线模型,除 MCS-Attention 外的注意力模块嵌入后均导致模型性能不同程度下滑。

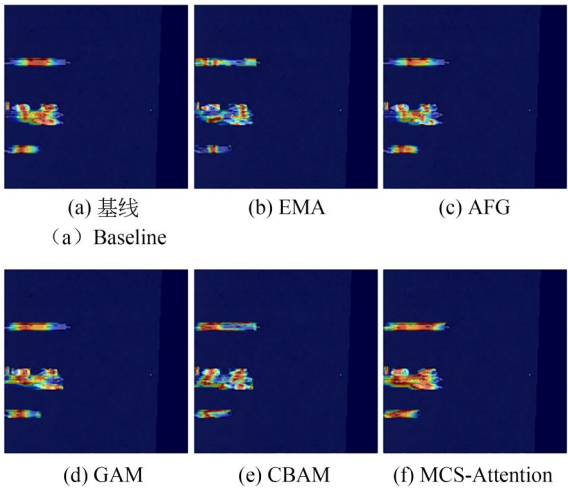


图 7 不同注意力机制的对比热力图

Fig. 7 Comparison heatmap of different attention mechanisms

4.3.3 损失函数对比实验

为验证 EMASlideLoss 的实质优势,将其与传统类均衡损失(Weighted BCE Loss)及经典难

样本优化损失(Focal Loss)进行定量对比。结果表明,EMASlideLoss 在 AP_{50} , AP_s , AP_m 指标上均优于两种传统方法,尤其在小目标分割上精度提升 1.5%,较 Focal Loss 提升 0.9%。相较于传统方法的固定权重固定衰减因子设计,

EMASlideLoss 通过时序窗口动态加权与指数移动平均平滑的双重机制,能实时适配遥感图像动态变化的类别分布,在不增加模型参数量与计算量的前提下,更高效地缓解类别不平衡带来的优化偏差。

表 4 损失函数对比实验结果

Tab. 4 Comparative experimental results of various dual-domain modules

Model	Params/M	AP_{50}	AP_s	AP_m	AP_l
基线	3.26	61.4	42.8	45.9	50.5
+Weighted BCE Loss ^[35]	3.26	62.1	43.5	46.8	51.0
+Focal Loss ^[36]	3.26	62.5	44.1	47.2	51.3
+EMASlideLoss	3.26	63.0	45.0	48.0	51.5

4.3.4 消融实验

为验证本文四个改进点在遥感图像实例分割中的改进效果,在 ISAID 数据集上设计了基于基线模型的五组消融实验,重点评估各改进模

块的性能增益。实验结果见表 5。其中,1,2,3和 4 分别表示在基线模型中依次加入 LGPM Block, MCS-Attention, SESS 和 EMASlideLoss 模块得到的模型。

表 5 消融实验结果

Tab. 5 Ablation experience

Model	Params/M	AP_{50}	AP_s	AP_m	AP_l	FPS
基线	3.26	61.4	42.8	45.9	50.5	104
+1	3.93	63.6	43.4	49.0	51.6	112
+3	3.58	62.8	47.5	48.2	51.1	98
+2+3	4.57	64.7	46.1	50.5	52.8	116
+2+4	4.01	62.0	44.6	47.2	51.1	109
+4+3+2+1	4.71	65.2	47.3	51.8	53.4	121

基于表 5 实验数据,基线模型在引入改进措施后,遥感图像实例分割性能呈现显著提升趋势。消融实验结果表明:单独引入 LGPM Block 模块可有效增强模型特征提取能力,促使 AP_{50} , AP_s , AP_m 及 AP_l 指标分别提升 2.2%, 0.6%, 3.1% 和 1.1%,推理速度从 104 FPS 提升至 112 FPS,在强化目标全局语义表征的同时,通过频率分支的傅里叶变换优化计算效率,为后续特征处理提供更丰富语义支撑;进一步集成 MCS-Attention 和 SESS 模块后,模型在多尺度目标分割任务中的表现显著优化, AP_{50} , AP_s , AP_m 及 AP_l 分别提升 0.6%, 3.3%, 1.3% 和 2.3%, MCS-Attention 结合卷积操作的局部特征提取能力与自注意力机制的上下文相关性构建能力,提升多

尺度目标分割精度,单独引入 SESS 模块时,模型对小目标的分割能力实现突破性提升: AP_s 从基线的 42.8% 提升至 47.5%,充分验证 SESS 模块通过四路径选择性扫描构建状态空间的有效性,可精准捕捉遥感小目标的边缘细节,避免小目标特征被复杂背景掩盖,推理速度保持 98 FPS;当上述四项改进策略集成应用时,模型综合性能达到最优, AP_{50} , AP_s , AP_m 及 AP_l 分别达到 65.2%, 47.3%, 51.8% 和 53.4%,各尺度 AP 指标均实现显著提升。研究结果验证了多模块优化策略在复杂遥感场景下对提升实例分割精度的有效性,特别是在小目标检测与多尺度特征融合方面的优势。

如图 8 所示的实例分割热力图结果显示,单

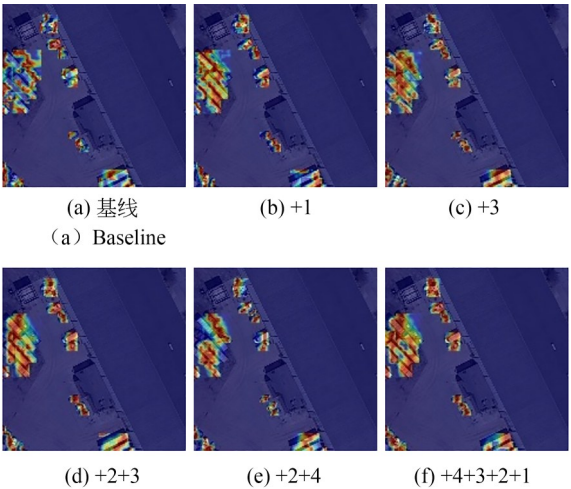


图8 消融实验热力图

Fig. 8 Ablation experience heatmap

独引入 SESS 模块后,小目标区域热力图激活可精准覆盖目标轮廓,激活完整性显著提升,以其四路径选择性扫描与低秩近似瓶颈卷积设计,有效强化了小目标特征捕捉。MCS-Attention 与 SESS 协同时,既保留小目标强激活特性,又能强化重叠区域语义区分,避免激活融合导致的定位偏差。LGPM Block 通过空间频率双域融合,显著抑制复杂背景杂波干扰。对比实验结果表明,在遥感图像实例分割场景中,本文模型生成的热

力图红色响应区域对远距离微小地物目标的聚焦度显著提升,且能更精准地界定地物目标与复杂背景的边界。这一特征充分印证了本文改进模型在遥感细节特征解析等方面具有更优异的性能表现。

4. 3. 5 在 ISAID 数据集的对比实验

为验证改进算法的实例分割性能,本研究在 ISAID 数据集上与主流实例分割算法开展对比实验。如表 6 所示,本文算法对 YOLOv8 模型的实例分割性能实现了显著优化:在效率层面,其参数量与浮点运算量仅略高于基线模型,避免了因性能提升导致的效率大幅损耗;同时,这两项指标显著低于 Mask R-CNN 系列,HTC, YOLACT 等其他主流算法,凸显出轻量化优势。在精度层面,该算法 AP_{50} 相较于 Box2MaskC 和 SOLO 对比算法,分别实现 14.6% 和 20.3% 的提升;且在多尺度目标分割任务中表现均衡,在 AP_s , AP_m 及 AP_l 的分割精度分别达到 47.3%, 51.8% 和 53.4%。全面覆盖遥感图像中不同尺度地物的分割需求。综上,本文所提模型不仅在总体检测精度上具备最优表现,其参数量与计算复杂度还明显低于主流同类模型;通过在检测精度与模型规模间实现更优的平衡,有效支撑了在边缘端设备的实际部署。

表 6 不同模型在 ISAID 数据集上的对比实验结果

Tab. 6 Comparative experimental results of different models on ISAID dataset

Model	Params/M	GFLOPs	AP_{50}	AP_s	AP_m	AP_l
Box2MaskC ^[37]	—	—	50.6	10.6	10.6	47.4
YOLACT ^[38]	34.83	28.5	43.3	8.4	31.8	40.4
SOLO ^[39]	—	—	44.9	24.2	7.9	32.7
Ref. ^[40]	43.83	32.1	54.5	15.5	37.8	42.0
HTC ^[41]	79.76	41.3	60.2	43.5	44.6	53.5
MaskR-CNN ^[42]	43.82	35.6	57.4	20.5	43.2	50.3
MaskScoring R-CNN ^[43]	60.09	38.7	57.7	20.8	44.3	51.5
ISDNet ^[44]	—	22.8	62.1	46.1	50.9	19.8
YOLOv8n ^[14]	3.26	8.1	61.4	45.8	48.9	50.5
SFE-YOLO ^[46]	7.12	9.8	43.5	38.8	21.2	43.1
PANet ^[47]	6.88	16.5	53.1	40.3	24.9	22.8
SPSIS ^[48]	5.36	23.2	25.3	39.2	32.8	34.7
MFE-Net ^[49]	9.26	15.1	51.4	35.8	47.9	40.5
本文	4.71	9.5	65.2	47.3	50.8	53.4

为直观验证算法改进的有效性,对部分经典算法与本文算法在 ISAIID 遥感图像舰船实例分割任务中的表现进行对比。

如图 9 所示的实例分割可视结果对比表明,在密集目标检测场景中,ISDNet 对密集排列的小尺寸易出现边缘融合现象,导致实例掩码重叠;MaskScoring R-CNN 对密集建筑群的定位偏移率较高,出现部分目标重复标注。YOLOv8n 虽能区分多数密集目标,但面对小目标仍存在漏检,反映其对密集场景下特征串扰的抑制能力不足。相比之下,本文模型通过 MCS-Attention 的局部自注意力机制与 SESS 的多路径扫描协同,有效强化了密集目标的边界语义区分,实现了掩码的精准分离,避免了重叠与漏检。在多尺度检测场景中,ISDNet 对小目标的分割完整性不足,掩码边缘残缺;MaskScoring R-CNN 对大尺寸飞机无法精准捕捉边缘等细微结构;YOLOv8n 虽

能兼顾中大型目标,但小目标漏检率较高。本文模型借助 LGPM Block 的空频双域特征融合与 SESS 的小目标特征增强,既实现了小目标掩码的完整覆盖,又保留了大目标的细节纹理,展现出优异的多尺度适配能力。在复杂环境检测场景中,ISDNet 和 MaskScoring R-CNN 易受背景杂波干扰:前者对植被遮挡下的目标出现误检,后者在阴影区域的目标检测精度不足;YOLOv8n 对低对比度目标的特征响应较弱,导致部分被遮挡目标漏检。本文模型通过 LGPM Block 的频率域傅里叶变换解耦背景干扰,结合 EMASlideLoss 动态调整难分样本权重,有效抑制了植被、阴影的影响,生成的掩码在复杂背景下仍保持较高完整性与准确性。综上,本文模型在遥感图像三类复杂场景中均表现出更优的分割鲁棒性与精度,验证了所提改进策略的有效性。

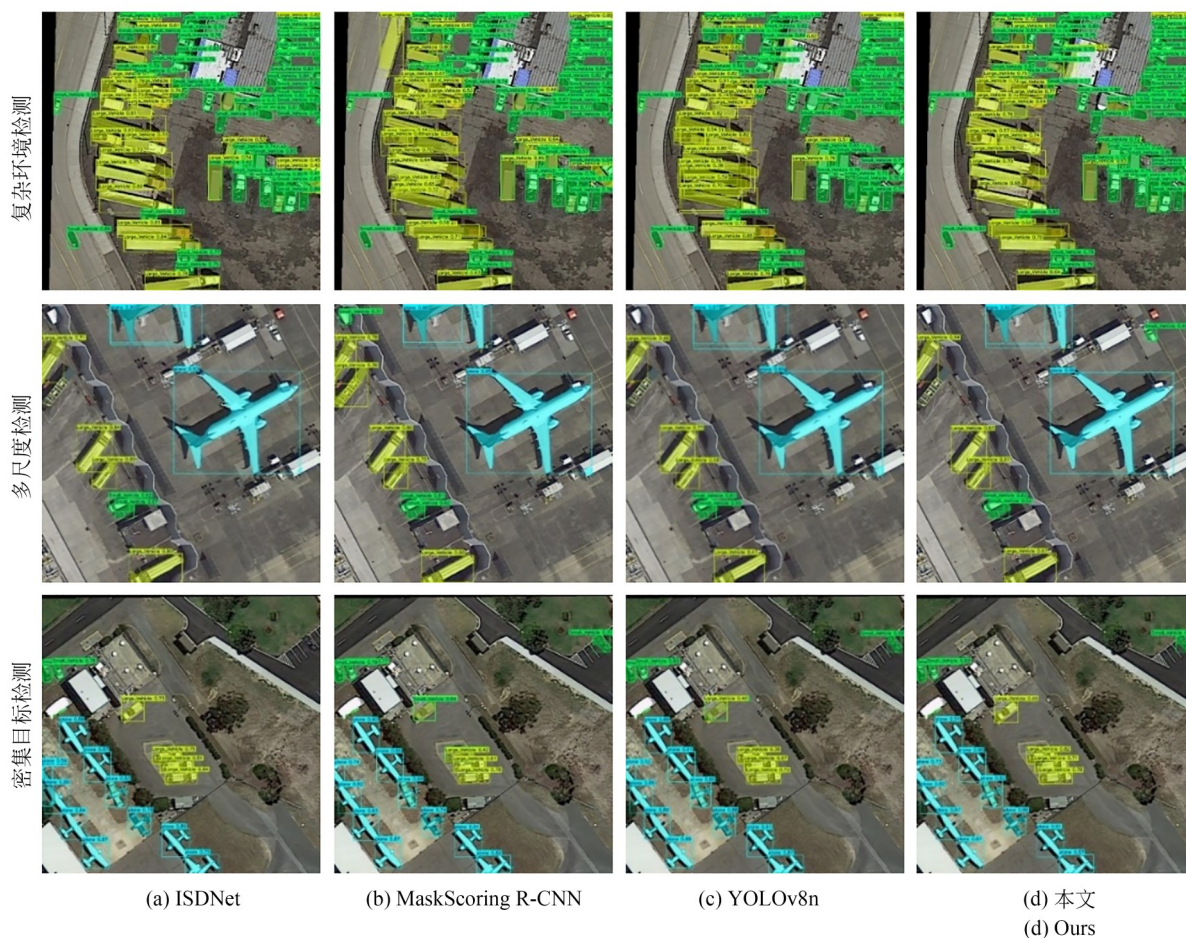


图 9 实例分割可视化结果对比

Fig. 9 Comparison of instance segmentation visualization results

4.3.6 在 NWPU VHR-10 数据集的对比实验

为进一步验证改进算法在遥感图像实例分割任务中的先进性,本研究在 NWPU VHR-10 数据集上开展横向对比实验,结果见表 7。实验结果显示:经典网络模型如 YOLACT, Mask R-CNN 及 R-CNN 等受限于特征提取架构,未能有

效解决多尺度目标特征融合问题,在复杂遥感场景下的实例分割精度提升有限。综合两组对比实验数据,本文改进算法在保持较低参数量实现了遥感图像实例分割精度的显著提升,通过多模块协同优化,有效强化复杂遥感场景下的特征表达能力。

表 7 不同模型在 NWPU VHR-10 数据集上的对比实验结果

Tab. 7 Comparative experimental results of different models on NWPU VHR-10 dataset

Model	AP_{50}	AP_s	AP_m	AP_l
YOLACT ^[38]	45.2	18.9	44.0	63.1
MaskR-CNN ^[42]	54.9	33.7	49.0	51.6
MaskScoring R-CNN ^[43]	57.0	46.7	56.9	55.8
HRNetV2p-W32 ^[45]	65.1	49.5	64.7	69.8
Ref. ^[40]	67.8	67.7	69.1	58.9
SFE-YOLO ^[46]	43.5	46.8	51.2	43.1
PANet ^[47]	61.8	44.5	59.7	61.5
SPSIS ^[48]	55.3	49.2	43.8	54.7
MFE-Net ^[49]	63.1	47.3	60.9	62.8
本文	68.3	50.7	68.1	64.2

5 结 论

针对遥感图像实例分割中的复杂目标表征与多尺度适配挑战,本文提出四点改进:构建局部—全局感知混频器,通过空间域多尺度卷积与频率域傅里叶变换的双分支融合,结合位置编码实现频率特征的差异化加权,提升重叠遮挡场景的语义感知能力;设计混合卷积自注意力模块,以卷积与自注意力并行结构平衡局部特征提取与全局上下文建模,通过动态权重筛选强化多尺度地物的特征聚合;提出结构增强状态空间模块,利用多路径扫描机制增强低对比度小目标的特征捕获能

力;引入指数移动平均滑动损失函数,通过时序窗口动态调整样本权重,结合动量机制缓解类别不平衡问题。本文模型相比基线模型 AP_{50} , AP_s , AP_m 及 AP_l 分别提升 3.8%, 1.5%, 1.9% 及 2.9%。各模块通过特征增强、多尺度建模、小目标优化及损失函数改进的协同作用,为高分辨率遥感影像提供高效实例分割方案。

作者贡献声明:

周冬梅:研究指导和论文审核;

贾琴琴:融合 Mamba 的多尺度遥感图像实例分割网络方法的提出,论文构思和撰写。

参考文献:

[1] GAO J L, ZHANG B, WU Y T, *et al.* Building Extraction from High Resolution Remote Sensing Images Based on Improved Mask R-CNN [C]. 2022 4th International Conference on Robotics and Computer Vision (ICRCV). September 25-27, 2022, Wuhan, China. IEEE, 2022: 1-6.

[2] YU D W, JI S P, LI X, *et al.* Earthquake crack de-

tection from aerial images using a deformable convolutional neural network [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 4412012.

[3] ZHU M M, HU G P, LI S, *et al.* A novel anchor-free method based on FCOS + ATSS for ship detection in SAR images [J]. *Remote Sensing*, 2022, 14(9): 2034.

[4] CHEN S L, OGAWA Y, ZHAO C B, *et al.*

- Large-scale individual building extraction from open-source satellite imagery *via* super-resolution-based instance segmentation approach[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 195: 129-152.
- [5] 何代毅,施文灶,林志斌,等. 基于改进 Mask-RCNN 的遥感影像建筑物提取[J]. 计算机系统应用, 2020, 29(9): 156-163.
- HE D Y, SHI W Z, LIN Z B, *et al.* Building extraction from remote sensing image based on improved mask-RCNN[J]. *Computer Systems & Applications*, 2020, 29(9): 156-163. (in Chinese)
- [6] WEI S J, ZENG X F, ZHANG H, *et al.* LFG-net: low-level feature guided network for precise ship instance segmentation in SAR images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5231017.
- [7] YASIR M, ZHAN L L, LIU S W, *et al.* Instance segmentation ship detection based on improved YoloV7 using complex background SAR images [J]. *Frontiers in Marine Science*, 2023, 10: 1113669.
- [8] HU M J, FENG D J, LI Q. Automatic extraction of buildings based on instance segmentation model [J]. *Bulletin of Surveying and Mapping*, 2020(4): 16-20.
- [9] 方俞泽,宋亚男,徐荣华. 基于轻量化 YOLACT 的遥感图像检测研究[J]. 无线互联科技, 2023, 20(10): 117-121.
- FANG Y Z, SONG Y N, XU R H. Research on Remote Sensing Image Detection Based on Lightweight YOLACT [J]. *Wireless Internet Technology*, 2023, 20(10): 117-121.
- [10] WANG X, ZHANG R, KONG T, *et al.* SO-LOv2: Dynamic and fast instance segmentation [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 17721-17732.
- [11] JIAO J B, LIU Y, LIU Y F, *et al.* VMamba: visual state space model[C]. *Advances in Neural Information Processing Systems* 37. December 10-15, 2024. Vancouver, BC, Canada. *Neural Information Processing Systems Foundation, Inc. (NeurIPS)*, 2024.
- [12] ZHUL, LIAOB, ZHANGQ, *et al.* Vision mamba: efficient visual representation learning with bidirectional state space model [C]. *Proceedings of the International Conference on Machine Learning*. Jul. 2024: 62429-62442.
- [13] HUANG T, PEI X, YOU S, *et al.* LocalMamba: Visual State Space Model with Windowed Selective Scan [EB/OL]. *arXiv preprint arXiv: 2024. 2403. 09338*.
- [14] VARGHESE R, M S. YOLOv8: a Novel Object Detection algorithm with enhanced performance and robustness[C]. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. April 18-19, 2024, Chennai, India. IEEE, 2024: 1-6.
- [15] MAO M, ZHANG R, ZHENG H, *et al.* Dual-stream network for visual recognition [J]. *arXiv preprint arXiv: 2105. 14734*, 2021.
- [16] LI X, WANG W, WU L, *et al.* Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. *arXiv preprint arXiv: 2006. 04388*, 2020.
- [17] DR A U R. Binary cross entropy with deep learning technique for Image classification [J]. *International Journal of Advanced Trends in Computer Science and Engineering*, 2020, 9(4): 5393-5397.
- [18] JIANG X Y, ZHANG X H, GAO N, *et al.* *When Fast Fourier Transform Meets Transformer for Image Restoration* [M]. Computer Vision-EC-CV 2024. Cham: Springer Nature Switzerland, 2024: 381-402.
- [19] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]. *Advances in Neural Information Processing Systems* 30. Red Hook: Curran Associates, 2017.
- [20] ZHANG X Y, ZHOU X Y, LIN M X, *et al.* ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 6848-6856.
- [21] 杨路,刘俊杰,余翔. 多尺度信息增强的遥感图像目标检测算法[J/OL]. 计算机工程, 2024: 1-13. (2024-11-21). [http://kns.cnki.net/KCMS/detail/detail.aspx? filename=JSJC20241120001&dbname=CJFD&dbcode=CJFQ](http://kns.cnki.net/KCMS/detail/detail.aspx?filename=JSJC20241120001&dbname=CJFD&dbcode=CJFQ).
YANG L, LIU J J, YU X. Multi-scale information Enhancement Algorithm for Remote Sensing Image Target Detection[J/OL]. *China Industrial Economics*, 2024: 1-13. (2024-11-21). <http://kns.cnki.net/KCMS/detail/detail.aspx? filename=JSJC20241120001&dbname=CJFD&dbcode=CJFQ>

- code=CJFQ. (in Chinese)
- [22] YANG C H, CHEN Z H, ESPINOSA M, *et al.* PlainMamba: Improving Non-Hierarchical Mamba in Visual Recognition [EB/OL]. *arXiv preprint arXiv*: 2024. 2403. 17695.
 - [23] LI J L, NIE Q, FU W F, *et al.* LORS: Low-rank residual structure for parameter-efficient network stacking [C]. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 15866-15876.
 - [24] LIU H, JIA C, SHI F, *et al.* Staircase Cascaded Fusion Of Lightweight Local Pattern Recognition and Long-Range Dependencies for Structural Crack Segmentation [EB/OL]. *arXiv preprint arXiv*: 2024. 2408. 12815.
 - [25] DING B L, ZHANG Y H, MA S. A lightweight real-time infrared object detection model based on YOLOv8 for unmanned aerial vehicles [J]. *Drones*, 2024, 8(9): 479.
 - [26] ZAMIR W S, ARORA A, GUPTA A, *et al.* iSAID: a large-scale dataset for instance segmentation in aerial images[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Long Beach, CA: IEEE, 2019: 28-37.
 - [27] 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.
 - [28] YANG J, QIU P J, ZHANG Y C, *et al.* D-Net: Dynamic large kernel with dynamic feature fusion for volumetric medical image segmentation [J]. *Biomedical Signal Processing and Control*, 2026, 113: 108837.
 - [29] MA X W, LIAN R R, WU Z K, *et al.* A novel scene coupling semantic mask network for remote sensing image segmentation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 221: 44-63.
 - [30] LI X, XU F, YU A Z, *et al.* A frequency decoupling network for semantic segmentation of remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5607921.
 - [31] 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 (ICASSP)*. June 4-10, 2023, Rhodes Island, Greece. IEEE, 2023: 1-5.
 - [32] SUN H, WEN Y, FENG H J, *et al.* Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing[J]. *Neural Networks*, 2024, 176: 106314.
 - [33] LIU Y, SHAO Z, HOFFMANN N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions [EB/OL]. *arXiv preprint arXiv*: 2021. 2112. 05561.
 - [34] WOO S, PARK J, LEE J Y, KWEON I S. CBAM: Convolutional Block Attention Module [EB/OL]. *arXiv preprint arXiv*: 2018. 1807. 06521.
 - [35] WEI J, WANG S, HUANG Q. FNet: Fusion, feedback and focus for salient object detection[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. New York: AAAI Press, 2020, 34 (7): 12321-12328.
 - [36] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection[C]. *2017 IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
 - [37] LI W, LIU W, ZHU J, *et al.* Box2Mask: box-supervised instance segmentation *via* level-set evolution[J]. *IEEE Trans Pattern Anal Mach Intell*, 2024, 46(7): 5157-5173.
 - [38] BOLYA D, ZHOU C, XIAO F Y, *et al.* YOLACT: Real-Time instance segmentation[C]. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019, Seoul, Korea (South). IEEE, 2019: 9156-9165.
 - [39] WANG X L, KONG T, SHEN C H, *et al.* SO-LO: Segmenting Objects by Locations[M]. *Computer Vision-ECCV 2020*. Cham: Springer International Publishing, 2020: 649-665.
 - [40] LUO Y X, HAN J M, LIU Z, *et al.* An elliptic centerness for object instance segmentation in aerial images [J]. *Journal of Remote Sensing*, 2022, 2022: 9809505.
 - [41] CHEN K, PANG J M, WANG J Q, *et al.* Hybrid task cascade for instance segmentation [C].

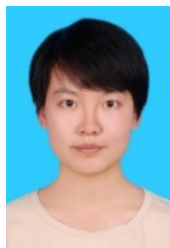
- 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 4969-4978.
- [42] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN[C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2980-2988.
- [43] HUANG Z J, HUANG L C, GONG Y C, *et al.* Mask scoring R-CNN[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 6402-6411.
- [44] GARG P, CHAKRAVARTHY A S, MANDAL M, *et al.* ISDNet: AI-enabled instance segmentation of aerial scenes for smart cities [J]. *ACM Transactions on Internet Technology*, 2021, 21 (3): 1-18.
- [45] KUMAR D. Accurate object detection & instance segmentation of remote sensing imagery using cascade mask R-CNN with HRNet backbone [J]. 2022: 1-12.
- [46] YU J Q, MAO P, WU W F, *et al.* TSNET: a solid waste instance segmentation model in China based on a Two-Step detection strategy and satellite remote sensing images [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2025, 136: 104366.
- [47] FU Y X, FANG W. A novel salient object detection network for burned area segmentation in high-resolution remote sensing images [J]. *Environmental Modelling & Software*, 2025, 193: 106629.
- [48] WANG Y D, PAN Y Z, LEI H, *et al.* SPSIS: single-point supervised instance segmentation for remote sensing [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 26267-26282.
- [49] LIU L, ZHANG S, HU M T. MFE-net: a novel multiscale feature enhancement network for SAR ship instance segmentation [J]. *IEEE Geoscience and Remote Sensing Letters*, 2025, 22: 1504305.

作者简介:



周冬梅(1975—),女,河南平顶山人,博士,副教授,硕士生导师,毕业于兰州交通大学,通信工程实验室主任,主要研究方向是无线网络、信号处理及人工智能。E-mail:912506338@qq.com

通讯作者:



贾琴琴(2000—),女,甘肃永靖人,硕士,研究方向:目标检测算法。E-mail:1657723517@qq.com