

文章编号 1004-924X(2026)14-2232-15

面向高分辨率遥感图像分割的级联精炼网络

刘春娟, 赵浩然, 闫浩文, 吴小所*, 包易航

(兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070)

摘要:针对高分辨率遥感图像语义分割中 Transformer 模型因多次下采样导致细节衰减、边缘模糊及结构性目标失真的问题,本文提出一种面向高分辨率遥感影像的级联精炼网络。该网络以 Swin-Transformer 为骨干编码器,在解码端依次构建了对偶特征聚合模块、动态小波融合上采样模块和几何结构感知模块。对偶特征聚合模块通过通道与空间协同建模增强多层特征互补,动态小波融合上采样模块利用多小波基自适应融合实现频域重建与细节恢复,几何结构感知模块则结合图结构先验与多尺度空间建模对末端特征进行几何校准,从而提升边界清晰度与目标完整性。在 ISPRS Potsdam, ISPRS Vaihingen 和 LoveDA 数据集上,该网络的平均交并比分别达到 87.13%, 83.86% 和 54.26%。该方法在复杂建筑轮廓刻画、小目标保持及跨场景分割任务中展现良好性能,验证了级联解码精炼策略在缓解特征逐层退化、改善遥感图像分割质量方面的有效性。

关键词:语义分割;遥感;小波变换;特征提取;边缘检测

中图分类号:TP751;TP391.41 **文献标识码:**A

doi:10.37188/OPE.20263414.2232 **CSTR:**32169.14.OPE.20263414.2232

Cascaded refinement network for high-resolution remote sensing image segmentation

LIU Chunjuan, ZHAO Haoran, YAN Haowen, WU Xiaosuo*, BAO Yihang

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

* Corresponding author, E-mail: wuxs_laser@lztu.edu.cn

Abstract: To address the issues of detail degradation, edge blurring, and structural target distortion arising from repeated downsampling operations in Transformer-based models for high-resolution remote sensing image semantic segmentation, this paper proposed a cascaded refinement network for high-resolution remote sensing imagery. The network employed Swin-Transformer as the backbone encoder and sequentially constructed three core modules on the decoder side, namely the dual feature aggregation module, the dynamic wavelet fusion upsampling module, and the geometric structure perception module. The dual feature aggregation module enhanced the semantic complementarity across multiple encoder layers through joint channel-spatial collaborative modeling. The dynamic wavelet fusion upsampling module performed frequency-domain reconstruction and fine-grained detail recovery via adaptive multi-wavelet basis fusion.

收稿日期:2026-04-23;**修订日期:**2026-06-17.

基金项目:国家重点研发计划(No. 2022YFB3903604);甘肃省自然科学基金资助项目(No. 21JR7RA310);青海理工学院“昆仑英才”人才引进科研项目(No. W2023-QLGKLYCZX-034)

The geometric structure perception module refined the terminal features through geometric calibration by integrating graph-structural priors with multi-scale spatial modeling, thereby improving boundary sharpness and target integrity. Extensive experiments on three public remote sensing benchmarks, including the ISPRS Potsdam dataset, the ISPRS Vaihingen dataset, and the LoveDA dataset, demonstrate that the proposed network achieves mean Intersection over Union scores of 87.13%, 83.86%, and 54.26%, respectively. The method exhibits superior performance in delineating complex building contours, preserving small-scale targets, and handling cross-scene segmentation tasks, which validates the effectiveness of the cascaded decoding refinement strategy in mitigating progressive feature degradation and enhancing the segmentation quality of remote sensing imagery.

Key words: semantic segmentation; remote sensing; wavelet transforms; feature extraction; edge detection

1 引 言

随着航空遥感技术发展,高分辨率影像^[1-2]语义分割已成为对地观测解译的核心手段。然而,复杂场景下地物尺度、纹理与边界差异显著,导致现有模型易出现边界模糊、结构破碎和小目标遗漏,细节信息与几何结构的稳定恢复仍是关键挑战。

现有的FCN^[3],DeepLab系列^[4]及U-Net^[5]等卷积神经网络虽能捕捉局部纹理,但在高分辨率遥感影像中感受野有限,难以有效融合多尺度特征,导致跨层融合不足、频域信息衰减与几何拓扑约束薄弱。Transformer模型虽具备全局语义建模优势,但在遥感图像分割中仍面临三个问题:浅层细节与深层语义之间的互补不足;高频边界与纹理信息缺少显式建模;结构性目标的拓扑关系约束不强。因此,复杂场景下的细节恢复与结构保持仍需进一步改进。

针对上述问题,已有工作分别从跨层融合、频域建模和几何拓扑约束等方向进行改进。跨层融合方面,为缓解编码-解码间语义差异与背景噪声干扰,注意力机制从早期的通道重校准^[6]演进至交叉注意力级联融合^[7]与多尺度选择性上下文^[8-9],通过跨尺度自适应重加权维持语义一致性。频域建模方面,为应对下采样引发的高频信息丢失与混叠现象^[10-11],小波卷积^[12]与频域Transformer^[13-14]引入显式频域建模,用于补充下采样过程中丢失的高频信息。几何拓扑约束方面,为纠正像素级分类引发的道路断裂、边界扭曲等形态失真,除常规边界增强^[15]外,形状先验网络^[16]及动态蛇形卷积^[17]通过显式建模像素间

关系,改善道路、建筑等结构性目标的形态表达。

然而,现有方法虽分别针对单一维度优化,却尚未充分考虑Transformer编解码过程中语义差异、细节衰减与结构失真之间的连续关联。因此,本文提出一种面向高分辨率遥感图像分割的级联精炼网络CR-Net。该网络以Swin-Transformer为编码器,在解码阶段构建“特征融合—细节恢复—几何校准”的三级级联优化框架,从而逐步缓解高分辨率遥感图像分割中的边界模糊、小目标遗漏和结构性目标轮廓失真问题。本文具体工作如下:

(1)针对编码器浅层细节特征与解码器深层语义特征之间的跨层语义差异,提出对偶特征聚合模块(Dual Feature Aggregation Module, DFAM)。该模块用于在通道与空间维度上融合编码器与解码器特征,实现浅层纹理与深层语义的自适应筛选和融合;

(2)针对多次下采样造成的高频边界信息衰减与空间细节损失,提出动态小波融合上采样模块(Dynamic Wavelet Fusion Upsampling, DW-FU)。该模块将传统空间域上采样转化为频域分解与重建过程,利用多小波基自适应加权机制联合恢复低频结构信息与高频细节响应;

(3)针对建筑、道路等结构性目标易出现的边界断裂、轮廓畸变和几何不一致问题,提出几何结构感知模块(Geometric Structure Perception Module, GSPM)。该模块在分割预测前引入图结构先验与多尺度空间建模,对像素邻域间的结构依赖关系进行显式建模,并对高分辨率特征进行几何校准。

2 CR-Net 模型

本文提出的级联精炼网络 CR-Net 整体架构如图 1 所示。网络采用 Swin-Transformer 作为编码器,由 Swin Block 1—Swin Block 4 逐级提取多尺度特征;解码器由对偶特征聚合模块 DFAM、动态小波融合上采样模块 DWFU 和几何结构感知模块 GSPM 组成。在解码过程中,编码器各层特征与对应解码特征依次通过 DFAM 实现跨层融合,随后经 DWFU 完成上采样与频域重建;同时,深层特征(F_4, F_3)通过交错式跨层残差路径注入浅层(F_2, F_1),以增强语义表达。最终,最高分辨率特征 F_1 在进入分割头前送入 GSPM 进行几何结构校准,输出最终分割结果。

2.1 对偶特征聚合模块的构建

Swin-Transformer 编码器虽能提取多尺度语义特征,但多级下采样会削弱局部纹理、细粒度边界和小尺度目标响应。针对传统拼接、相加等直接融合方式易引入背景噪声、造成边界模糊和类间混淆的问题,本文提出对偶特征聚合模块 DFAM,不同于单路注意力重标定,DFAM 以编码器特征与解码器特征的双路交互为核心,在通道和空间维度上对跨层信息进行自适应筛选。模块结构如图 2 所示。

在通道注意力分支中,为增强与目标判别相关的语义响应并抑制冗余背景通道,模块分别对编码器特征 F_{en} 与解码器特征 F_{de} 执行全局平均池化与最大池化,并将所得统计特征融合为联合

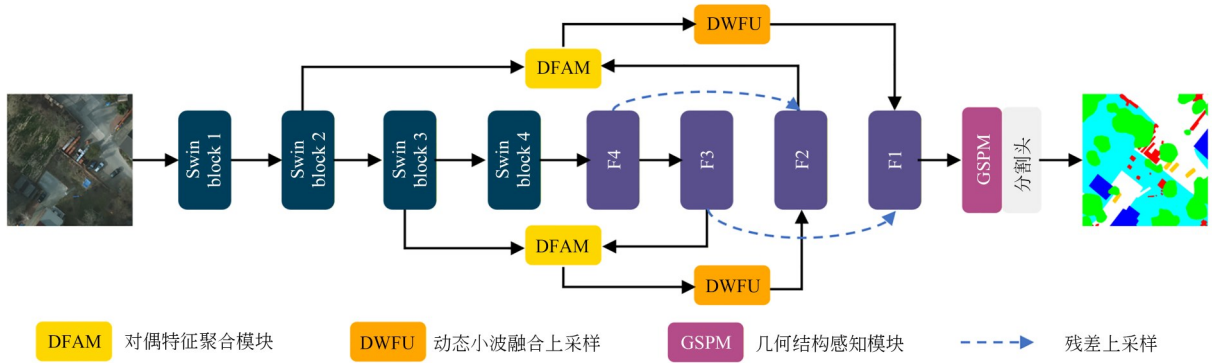


图 1 CR-Net 模型结构图

Fig. 1 CR-Net module structure diagram

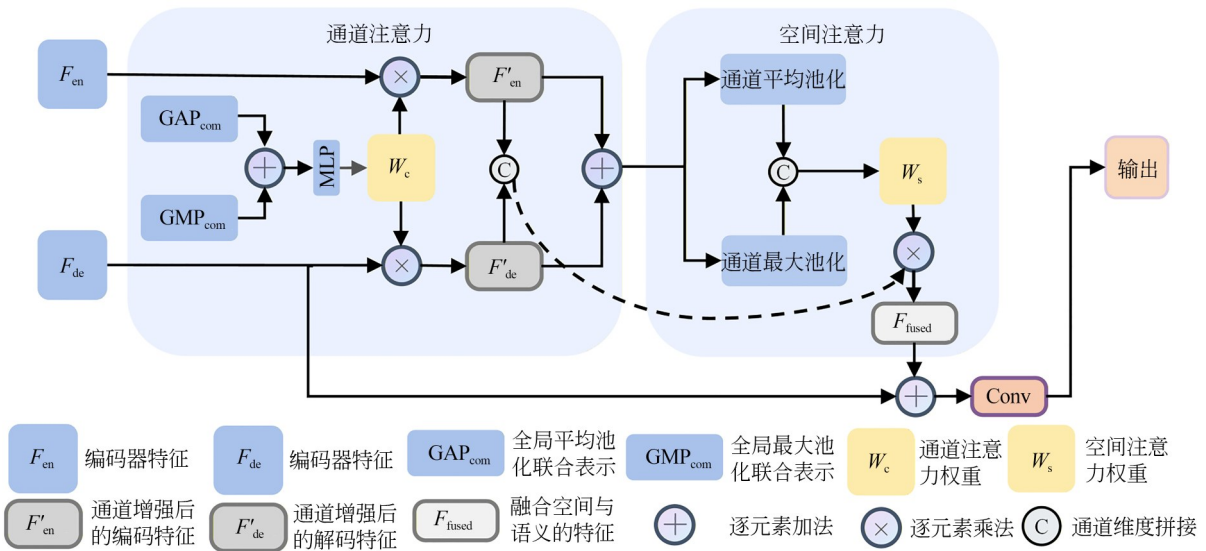


图 2 DFAM 模块结构图

Fig. 2 DFAM module structure diagram

表示 GAP_{com} 与 GMP_{com} ,以同时刻画整体响应分布和显著激活区域。如图 2 所示,融合后的特征经由两层 MLP 构成的共享门控网络,生成通道注意力权重 W_c ,如式(1)所示:

$$W_c = \sigma \left(W_1 \left(\text{ReLU} \left(W_0 (GAP_{com} + GMP_{com}) \right) \right) \right), \quad (1)$$

其中: W_1 和 W_0 为线性变换, σ 为 Sigmoid 函数。该权重分别与编码器特征 F_{en} 和解码器特征 F_{de} 进行元素级乘法交互,得到通道增强后的特征 F'_{en} 与 F'_{de} 。通过这一过程,模块对不同通道进行重标定,使与目标类别相关的响应得到保留,并减弱背景纹理、阴影和局部噪声的影响。

考虑到遥感图像中建筑边缘、车辆等目标仅占据局部区域,空间注意力分支进一步关注特征在空间维度上的关键位置。对通道增强后的两路特征 F'_{en} 与 F'_{de} 进行元素级相加后,构建包含平均池化与最大池化信息的空间描述图,并生成空间注意力权重 W_s ,该权重旨在引导网络更加关注边界过渡区域、小目标所在区域以及局部结构变化较明显的位置,从而有助于缓解简单融合时背景区域对结果的干扰。该过程如式(2)所示:

$$W_s = \sigma \left(\text{Conv} \left(\text{Concat} \left(\text{AvgPool}, \text{MaxPool} \right) \right) \right). \quad (2)$$

在此基础上,两路特征 F'_{en} 与 F'_{de} 在通道维度上拼接融合后,乘以空间注意力权重 W_s (图 2 中虚线所示)进行空间权重调制,形成融合空间信息与语义信息的融合特征 F_{fused} 。

完成通道与空间调制后,DFAM 将融合特征 F_{fused} 与原始解码器特征 F_{de} 进行残差式结合,并通过由 3×3 卷积、批归一化和 ReLU 组成的卷积块完成特征平滑与通道重映射,输出最终特征。该过程旨在减小编解码特征差异,并为后续上采样与分割预测提供更具判别性的输入,该过程如式(3)所示:

$$\text{Output} = \text{Conv}_{3 \times 3} \left(\text{ReLU} \left(\text{BN} \left(\text{Conv}_{3 \times 3} (X) \right) \right) \right). \quad (3)$$

综上,DFAM 在深层语义约束下筛选并重组浅层细节特征,生成跨层信息更协调、背景干扰较弱的融合特征,为 DWFU 频域重建提供稳定输入,并降低上采样过程中无关高频纹理被放大的风险。

2.2 动态小波融合上采样模块的构建

经过 DFAM 处理后,解码特征的跨层语义一致性得到改善,但其空间分辨率仍需逐级恢复。最近邻插值、双线性插值等常规上采样方法通常在空间域直接放大特征,难以显式区分低频结构信息与高频边界细节,容易在复杂地物边界处产生混叠、伪影和轮廓平滑问题。针对多级下采样导致的高频边界衰减和空间细节损失,本文提出了动态小波融合上采样模块 DWFU。该模块由两部分组成:小波变换用于频域分解,多小波基自适应上采样单元用于特征重构,旨在提升遥感图像中复杂边界和多尺度地物的重建精度。

为适应遥感图像中边界突变、纹理细节和平滑过渡区域并存的特点,DWFU 依据支撑长度、消失矩、对称性和正则性等性质,选取 Db1, Sym4 和 Db4 三类小波基构成互补的频域重建集合^[18-19]。Db1 即 Haar 小波,具有最短支撑和较强的局部突变响应能力,适合捕获建筑轮廓、道路边缘等高频边界信息;Sym4 为近似对称的 Symlet 小波,可在重建过程中减小相位偏移,有助于保持平滑区域和边界过渡区域的结构一致性;Db4 属于 Daubechies 正交紧支撑小波,相较 Db1 具有更高的平滑性和更丰富的频率表征能力,适合复杂纹理区域的细节恢复。三者共同构成互补的频域重建基。图 3 展示了 DWFU 模块的整体流程。

在频域分解阶段,DWFU 通过小波变换 (Wavelet Transform, WT) 将输入特征图分解为低频和高频分量,其中低频分量 X_{LL} 主要保留整体结构与粗粒度语义信息,高频分量 X_{LH} , X_{HL} 和 X_{HH} 分别对应不同方向的边缘、纹理和局部突变信息。因此,将上采样过程建立在频域分解基础上,有助于显式区分结构恢复与细节重建,减少单一空间域放大引起的信息混叠。

在高频分量处理阶段,模块引入压缩激励模块 (Squeeze-and-Excitation, SE) 对高频特征进行通道重构,并采用并行卷积分支处理不同方向的边界信息。这些分支分别以水平 Sobel 核、垂直 Sobel 核和 Laplacian 核进行初始化,用于定向生成对应的高频细节分量 X_{LH} , X_{HL} 和 X_{HH} , 并与低频分量共同堆叠为结构化子带张量 X_{sub} , 为后续逆小波重构提供输入。

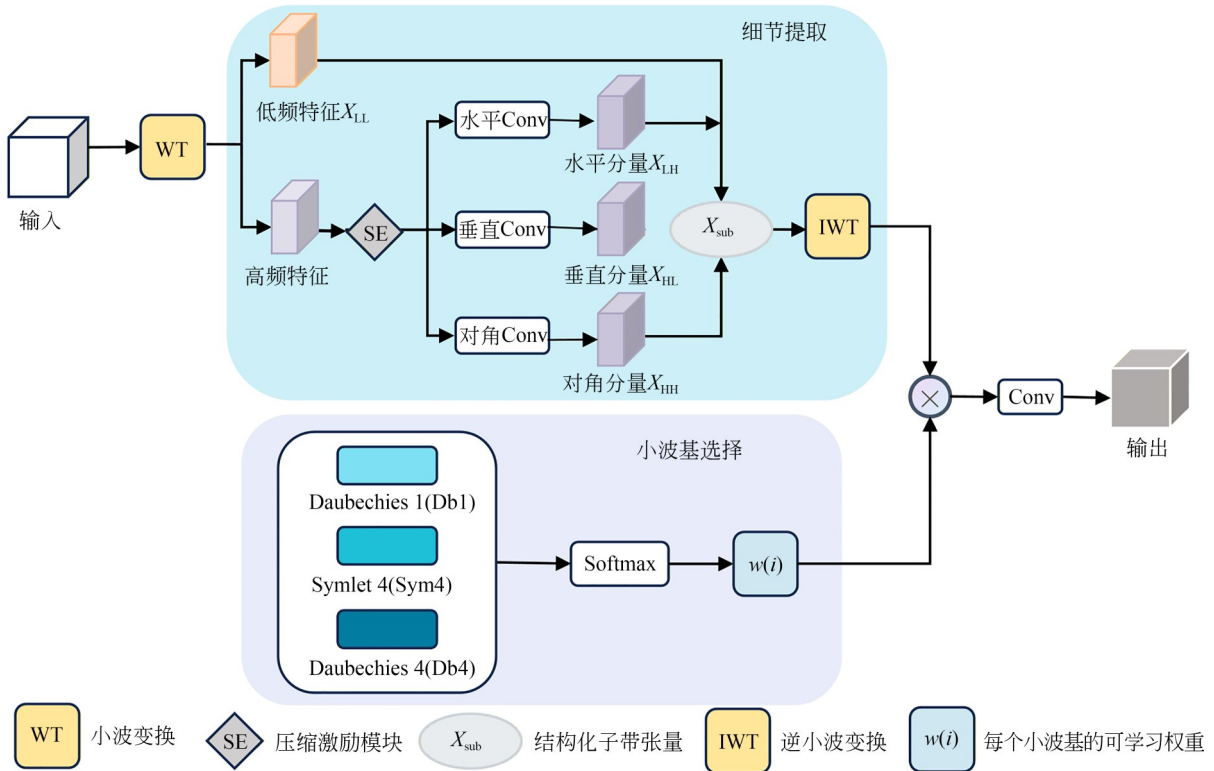


图3 DWFU模块结构图

Fig. 3 DWFU module structure diagram

在特征重建阶段,模块通过 Softmax 函数生成不同小波基的归一化权重 w_i ,对包含完整频率信息的子带 X_{sub} 执行逆小波变换 (Inverse Wavelet Transform, IWT),完成特征图从频域到空域的重建。基于 WT 与 IWT 的重建方式可分别建模低频结构与高频细节,缓解直接插值上采样带来的高频衰减和边缘平滑问题。融合过程如式 (4) 所示:

$$X_{up} = \sum_{i=1}^3 \text{softmax}(w_i) \cdot \text{IWT}(X_{sub}, F_{reg}^{(i)}), \quad (4)$$

其中: w_i 为每个小波基的可学习权重, $F_{reg}^{(i)}$ 为不同小波滤波器提供了不同的频率响应形式,用于适配遥感图像中的多种特征。完成空间重建后, DWFU 再通过卷积减轻混叠效应。

经过 DWFU 后,解码特征不仅完成分辨率恢复,而且在目标边界邻域获得更清晰的高频响应,为后续 GSPM 的图结构建模和几何校准提供更可靠的边缘基础。

2.3 几何结构感知模块的构建

经过 DFAM 与 DWFU 处理后,解码特征的跨层语义一致性和高频细节表达得到改善,但像

素级预测仍缺乏对建筑轮廓闭合、道路连通性及细长目标方向一致性的显式几何约束,易产生边界断裂、局部侵蚀和结构失真等问题。常规边界增强或边缘监督方法通常依赖额外边界标注或局部卷积感知,难以直接刻画像素邻域间的结构关系。因此,本文在分割头前引入几何结构感知模块 GSPM,对语义融合与频域重建后的高分辨率特征进行末端几何校准。该模块结合图结构先验与动态蛇形卷积,通过图特征提取、图特征聚合和空间注意力进行协同建模,如图 4 所示。

具体而言, GSPM 首先构建图顶点-边交互机制。输入特征通过深度可分离卷积生成顶点特征,同时采用四方向移位操作显示构造方向边,从而将像素间几何依赖关系表示为顶点表征与边表征的交互过程。顶点特征与边特征分别进入更新支路:顶点更新支路通过拼接与线性变换得到更新顶点,边更新支路通过拼接增强边与顶点上下文信息得到更新边。图 4 中的 Linear 层用于对拼接后的顶点一边联合表示进行通道维度投影,而非执行固定窗口内的空间卷积。由于 GSPM 已通过四方向移位显式建立局部图邻接

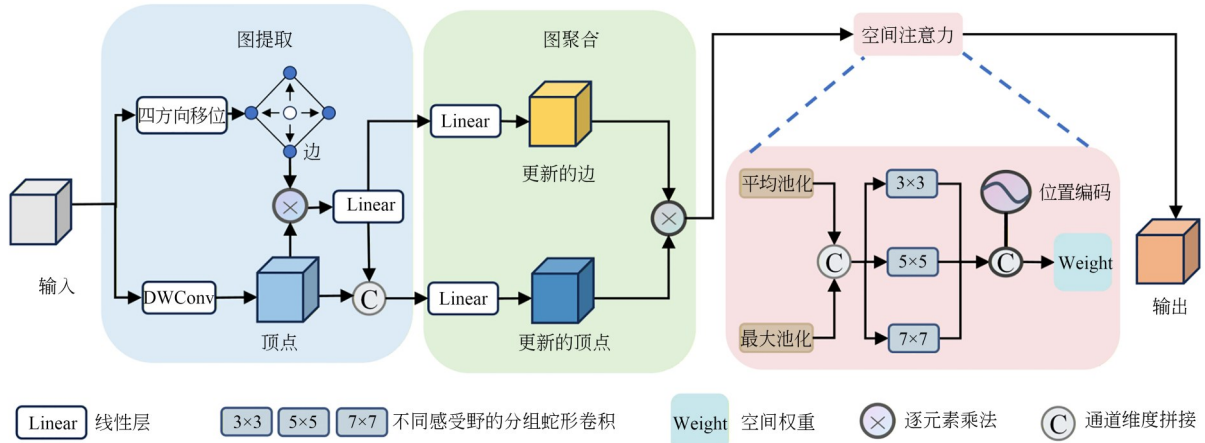


图4 GSPM模块结构图

Fig. 4 GSPM module structure diagram

关系,线性映射可在不额外引入卷积核空间约束的情况下完成点-边关系重组,使顶点更新和边更新更加轻量且具有更高的关系建模自由度。最后,更新后的边特征作为门控因子与更新顶点进行元素级交互,实现边缘结构信息对节点表达的约束。

考虑到遥感图像中同时存在车辆等微小目标和道路等大跨度连通目标,GSPM进一步引入多尺度空间注意力机制。输入特征经平均池化与最大池化压缩通道冗余后,被送入 3×3 、 5×5 和 7×7 三路并行分组蛇形卷积分支,分别提取局部边界、中尺度结构和大范围上下文特征 F_3 、 F_5 和 F_7 。为降低高分辨率特征上的计算开销,三路分支仅作用于图更新后的一半通道,另一半通道则用于保留原始几何响应,从而避免对全部通道进行冗余卷积。对于单个分支,其主要空间卷积复杂度由普通卷积的 $o(HW\cdot C^2\cdot K^2)$ 降低为 $o(HW\cdot C/2\cdot K^2)$,其中 H 、 W 、 C 分别表示特征图高度、宽度和通道数。随后,将多尺度特征在通道维度拼接,并与可学习位置编码融合,经卷积与Sigmoid函数生成空间权重Weight。该空间权重用于增强边界过渡和连通结构区域的响应,使几何精炼更加集中于结构易失真的位置,从而提升目标轮廓的连续性。其计算过程如式(5)所示:

$$Weight = \sigma \left(Conv_{1\times 1} \left(Concat(F_3, F_5, F_7) \right) + Pos_{enc} \right). \quad (5)$$

3 实验设计与结果分析

3.1 数据集

本文采用经典航空遥感数据集ISPRS Potsdam, ISPRS Vaihingen以及LoveDA开展实验验证。Potsdam与Vaihingen数据集均标注6个语义类别,包括不可渗透地表、建筑、低矮植被、树、车和杂物。Potsdam数据集包含38张 $6\,000\times 6\,000$ pixels, 5 cm分辨率图像,其中23张用于训练、15张用于测试。Vaihingen数据集包含33张约 $2\,000\times 4\,000$ pixels, 9 cm分辨率图像,其中16张用于训练、17张用于测试。LoveDA数据集由5 987张覆盖南京、常州和武汉三市的高空间分辨率图像组成,其中城市场景2 713张、农村场景3 274张。该数据集标注7个类别,包括背景、建筑、道路、水面、荒地、森林和农田,以评估模型在显著域间差异下的泛化能力。为保证输入尺度一致,所有图像均裁剪为 $1\,024\times 1\,024$ pixels用于训练。

3.2 损失函数

为兼顾类别不平衡场景下的训练稳定性与像素分类精度,本文采用由Dice损失与软交叉熵损失(Soft Cross-Entropy Loss)组成的联合损失函数。Dice损失用于优化区域重叠度,并处理前景与背景分布不均的问题;软交叉熵损失用于提供更平滑的像素级梯度。因此,本文将二者进行加权线性组合,构建互补型联合损失函数,以兼顾前景区域建模和像素级梯度稳定性,计算公式

如式(6)~式(8)所示:

$$I_{\text{dice}} = 1 - \frac{2}{N} \sum_{n=1}^N \sum_{k=1}^K \frac{y_k^{(n)} \hat{y}_k^{(n)}}{y_k^{(n)} + \hat{y}_k^{(n)}}, \quad (6)$$

$$I_{\text{ce}} = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^{(n)} \log y_k^{(n)}, \quad (7)$$

$$I_{\text{main}} = \alpha I_{\text{dice}} + (1 - \alpha) I_{\text{ce}}, \quad (8)$$

其中: N 代表样本数量, K 代表类别数。 $y_k^{(n)}$ 表示正确标签, $\hat{y}_k^{(n)}$ 表示预测结果,其中的 k 表示第 k 个类别, (n) 表示图中有 n 个样本。 $\alpha \in [0, 1]$,为Dice损失的权重。 $1 - \alpha$ 为软交叉熵损失的权重。

为验证不同权重比例的最优平衡,本文在Potsdam数据集上开展了关于Dice损失权重 α 的消融实验,如表1所示。实验结果表明,当 $\alpha = 0.5$,即Dice损失和软交叉熵损失等权线性组合时,可以取得87.13%的mIoU最佳性能,优于其他比例及其他单一损失项。这一结果说明等权组合能有效平衡两项损失的优势,避免单一损失的局限性。

需要说明的是,本文未额外引入显式边界损失项。边界模糊问题主要通过网络结构中的频域细节恢复与几何结构校准机制进行缓解,即由DWFU增强高频边界信息,由GSPM对解码末端特征进行结构一致性约束。该设计避免了额外边界标签生成和边界损失权重调节带来的训练复杂度,同时保持了实验设置的一致性。

表 1 不同损失权重在 Potsdam 数据集上的消融实验结果
Tab.1 Ablation results of different loss weights on the Potsdam dataset

α	$1 - \alpha$	mIoU
0	1.0	85.53
0.3	0.7	86.47
0.5	0.5	87.13
0.7	0.3	86.59
1.0	0	85.72

3.3 工具及评价指标

本文所有实验均在PyTorch 2.3.1框架下完成,硬件平台配置为NVIDIA GeForce RTX 3090 GPU (24GB 显存)及Intel Xeon E5-2680 v4 CPU。模型采用AdamW优化器^[20],训练批次固定为8,并设置早停25轮防止过拟合。骨干

网络采用在ImageNet-22K数据集上预训练的权重进行初始化。考虑到CR-Net的训练同时包含预训练Transformer编码器微调与解码端新增模块学习,初始学习率需兼顾表征稳定性和梯度更新有效性。过大的学习率可能破坏骨干网络的通用视觉表征并导致梯度波动,过小的学习率则会降低DFAM,DWFU和GSPM的收敛效率。基于预训练Transformer微调任务中常用的小学习率设置,并结合训练过程的稳定性需求,本文将初始学习率设为 $6e-5$,权重衰减设置为0.01。学习率调度采用带热重启的余弦退火策略,初始重启周期设为 $T_0 = 15$,周期倍增因子 $T_{\text{mult}} = 2$,以增强训练后期的优化稳定性和周期性再搜索能力。

为全面评估模型的分割性能,本研究采用整体准确率(Overall Accuracy, OA)、平均交并比(mean Intersection over Union, mIoU)以及平均F1分数(mean F1-Score, mF1)作为评价指标。各指标基于标准的混淆矩阵元素(TP, FP, FN, TN)如式(9)~式(13):

$$\text{precision}_k = \frac{TP_k}{TP_k + FP_k}, \quad (9)$$

$$\text{recall}_k = \frac{TP_k}{TP_k + FN_k}, \quad (10)$$

$$F1_k = 2 \times \frac{\text{precision}_k \times \text{recall}_k}{\text{precision}_k + \text{recall}_k}, \quad (11)$$

$$OA = \frac{\sum_{k=1}^K TP_k}{\sum_{k=1}^K TP_k + FP_k + TN_k + FN_k}, \quad (12)$$

$$\text{mIoU} = \frac{1}{K} \sum_{k=1}^K \frac{TP_k}{TP_k + FP_k + FN_k}, \quad (13)$$

其中: K 代表类别数,下标 k 表示第 k 个类别。F1分数是精确率(Precision)和召回率(Recall)的调和平均值,为每个类别提供平衡的性能度量;mIoU作为语义分割的关键指标,用于评估预测结果与真实掩码的空间重叠度;mF1和OA分别从类别均衡性和整体像素分类正确率两个角度补充分割结果。

3.4 消融实验

3.4.1 核心模块消融实验

为分析各核心组件的作用,本文在Potsdam数据集上以Swin-Tiny为基线模型开展消融实验。基线解码器采用空洞深度可分离卷积调整通道与感受野,随后通过最近邻插值恢复分辨

率;所有配置均保持编解码主干一致,仅替换或添加目标模块作为性能增益基准。在引入DWFU时,基线中的空洞卷积与插值操作被移除,由DWFU承担上采样过程。

表2结果表明,DFAM,DWFU和GSPM均具有独立有效性。相较于Baseline,三者分别使mIoU提升2.08%,2.17%和2.07%。其中,DWFU在单模块配置下取得最高精度,mIoU,

mF1和OA分别达到86.76%,92.79%和91.50%,且GFLOPs仅为29.50,表明频域重建能够以适度计算开销改善上采样细节恢复。GSPM以最低复杂度增量获得与DWFU相近的性能提升,其Params和GFLOPs较Baseline仅增加1.19M和1.62,体现出较高的参数与计算效率;DFAM则显著增强跨层特征融合能力,但引入的额外计算开销相对较高。

表 2 ISPRS Potsdam 数据集上的消融实验结果

Tab. 2 Ablation Experiment Results on the ISPRS Potsdam Dataset

方法	DFAM	DWFU	GSPM	Params/M	GFLOPs	mF1/%	OA/%	mIoU/%
Baseline				28.86	26.97	91.55	90.62	84.59
+DFAM	✓			31.19	30.64	92.73	91.28	86.67
+DWFU		✓		31.28	29.50	92.79	91.50	86.76
+GSPM			✓	30.05	28.59	92.73	91.28	86.66
+DFAM+DWFU	✓	✓		33.61	33.18	92.82	91.40	86.79
+DFAM+GSPM	✓		✓	32.38	32.26	92.79	91.39	86.76
+DWFU+GSPM		✓	✓	32.48	31.12	92.85	91.46	86.86
本文	✓	✓	✓	34.80	34.79	93.01	91.56	87.13

双模块组合的性能均优于其对应单模块配置,其中DWFU+GSPM取得最高86.86%的mIoU,且Params和GFLOPs分别为32.48M和31.12,表明频域细节恢复与几何结构校准具有较强的功能互补性。完整集成三个模块后,模型mIoU,mF1和OA分别达到87.13%,93.01%和91.56%,较Baseline提升2.54%,1.46%和0.94%;Params和GFLOPs分别增加5.94M和7.82。上述结果验证了级联精炼策略的有效性,即在可控的复杂度增量下,实现跨层特征融合、细节重建与几何校准的协同优化,并取得最优综合分割性能。

3.4.2 动态小波融合上采样小波基消融分析

如表3所示,各小波重构配置的mIoU均优于Baseline,表明将传统空间域上采样替换为频域重建能够有效改善解码特征恢复。单小波基配置中,Db1取得最高的85.91% mIoU,较Baseline提升1.32%,说明其短支撑与局部响应特性更适用于当前上采样重建任务;Sym4和Db4的mIoU分别为85.67%和85.39%,同样获得稳定增益,表明不同频率响应形式均可了解码特征重建提供有效补充。

表 3 Potsdam 数据集上 DWFU 小波基消融结果

Tab. 3 Ablation of wavelet bases in DWFU on the Potsdam dataset

小波基	Params/M	GFLOPs	mIoU
Baseline	28.86	26.97	84.59
Db1	31.14	29.20	85.91
Sym4	31.21	29.34	85.67
Db4	31.20	29.34	85.39
Db1+Sym4	31.21	29.35	86.48
Db1+Db4	31.21	29.35	86.31
Sym4+Db4	31.28	29.49	86.08
Db1+Sym4+Db4	31.28	29.50	86.76

多小波基组合进一步提升了分割性能。双小波基配置中,Db1+Sym4取得取得86.48%的最高mIoU,较Baseline提升1.89%;Db1+Db4和Sym4+Db4分别达到86.31%和86.08%,均优于对应单小波基配置,说明联合建模能够缓解单一小波基频率响应能力的局限。进一步融合Db1,Sym4和Db4后,模型取得最高的86.76% mIoU,较Baseline提升2.17%;Params和GFLOPs分别为31.28M和29.50,较Baseline仅增加2.42M和2.53GFLOPs。上述结果表明,多小波基自适

应融合能够以有限复杂度增量增强频域重建表征,使DWFU在解码上采样阶段获得更稳定的细节恢复能力。

3.4.3 几何结构感知模块多尺度卷积消融分析

如表4所示,3×3+5×5+7×7完整多尺度聚合取得最高的86.66% mIoU,表明不同感受野信息的联合建模有助于提升GSPM的特征精炼能力。单尺度中,3×3分支表现最佳,说明较小感受野在当前高分辨率末端特征处理中具有更好的适配性;双尺度中,3×3+5×5组合优于其他组合,表明局部响应与中尺度上下文的结合能够进一步增强特征表达。加入7×7分支后,模型可整合更大范围的上下文信息,因此获得最优分割性能。

同时,不同尺度组合仅带来较小的Params和GFLOPs增幅。这主要得益于GSPM中分组蛇形卷积的轻量化设计,以及多尺度分支仅作用于图更新后一半通道的结构设置。使其复杂度主要随 $(C/2) \times K^2$ 增长,而非普通卷积的 $C^2 \times K^2$ 。由此可见,多尺度卷积分支在几乎不显著增加模型复杂度的情况下,提升了GSPM对末端特征的多感受野建模能力,为最终结构校准提供了更充分的空间表征。

表4 Potsdam数据集上GSPM多尺度卷积消融结果

Tab.4 Ablation of multi-scale convolutions in GSPM on the Potsdam dataset

卷积尺度	Params/M	GFLOPs	mIoU
3×3	30.04	28.450	85.91
5×5	30.043	28.462	85.78
7×7	30.044	28.481	85.55
3×3+5×5	30.045	28.510	86.43
3×3+7×7	30.047	28.529	86.27
5×5+7×7	30.047	28.542	86.16
3×3+5×5+7×7	30.050	28.590	86.66

3.4.4 动态小波融合上采样可视化分析

为分析DWFU模块在不同阶段的作用,本文统计了两次上采样阶段的小波基权重,并对对应特征进行了可视化,结果如表5和图5所示。表5显示,两个上采样阶段中Db1均具有最高融合权重,且其权重由第一阶段的0.375上升至第二阶段的0.400,表明随着解码过程接近输出端,

模型对Db1重建分支的依赖有所增强。结合表3中Db1在单小波基配置下取得最高mIoU的结果,可以推测短支撑小波基更有利于解码后期的局部细节恢复。随着解码过程接近输出端,Db1权重上升,说明模型在后期分辨率恢复阶段对高频轮廓信息的利用有所增强。

图5展示了两次动态小波上采样前后的特征变化。经DWFU处理后,特征响应的空间区分度有所增强,为其改善上采样特征表达提供了定性支持。该现象与表1中第二次上采样阶段Db1权重升高的趋势一致,表明多小波基动态加权机制能够强化频域层面的边缘细节建模,并提升上采样特征的轮廓表达能力。

表5 两次动态小波上采样阶段的小波基权重结果

Tab.5 Wavelet basis weight results of two dynamic wavelet upsampling stages

	Db1	Sym4	Db4
第一次小波上采样	0.375	0.337	0.288
第二次小波上采样	0.400	0.341	0.259

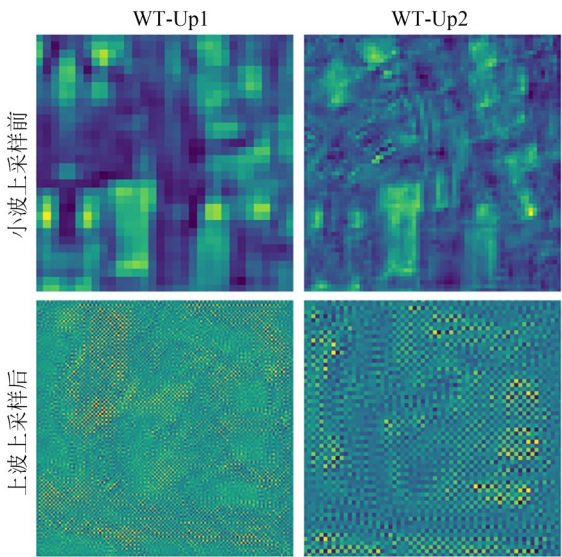


图5 两次动态小波上采样前后可视化结果

Fig.5 Before-and-after visualizations of two dynamic wavelet upsampling stages

3.5 对比实验

为全面评估CR-Net的性能,本文将其与多种代表性方法进行对比,涵盖多种主流架构,具体参数如表6所示。对比方法包括:基于卷积

神经网络的分割方法,如 CMTFNet, AFENet^[21]和 GDGNet^[22],用以作为局部细节建模能力的对照;基于 Transformer 及注意力机制的混合架构,如 UANet^[23],用于评估模型在全局语义建模与多尺度特征融合方面的表现;同时,引入 PyramidMamba^[24]和 RS3Mamba^[25]等近期方法,用于比较所提模型在不同类型方法中的实验结果。

表 6 网络参数对比

Tab. 6 Comparison of network parameters

网络	骨干网络	Params/M	GFLOPs
PyramidMamba	ResNet-18	24.40	18.71
GDGNet	ResNet-18	26.89	55.22
UANet	pvtv2-b2	25.64	55.37
CMTFNet	ResNet-50	30.07	68.02
AFENet	ResNet-18	20.24	53.53
RS3Mamba	R18, Mamba-T	43.32	31.65
CR-Net	Swin-T	34.80	34.79

3.5.1 在 ISPRS Potsdam 数据集上的对比实验

如表 7 所示,CR-Net 在 ISPRS Potsdam 数据集上取得了 87.13% 的 mIoU, 91.56% 的 OA 和 93.01% 的 mF1,表现处于先进水平。与 CMTF-Net 相比,CR-Net 的 mIoU, OA 和 mF1 分别提升了 0.10%, 0.03% 和 0.08%;与 PyramidMamba 相比,分别提升了 0.91%, 0.32% 和 0.59%,表

明所提方法在指标上保持较高的整体精度。

进一步从类别结果来看,CR-Net 在“低矮植被”、“树”和“车”类别上的 F1 分数分别达到 88.33%, 89.56% 和 96.72%。这些结果说明,该模型在相邻类别区分、复杂纹理表征和小目标保持方面具有一定作用。

图 6 的可视化结果与表 7 中的定量结果基本一致。如图中红框区域所示(彩图见期刊电子版),相比部分对比方法在建筑边缘处出现的锯齿和模糊现象,CR-Net 得到的分割掩膜边界更连续,阴影区域的误分也有所减少。

在地物分布密集的场景中(如第 4 行),CR-Net 的预测区域一致性相对更好。对于车辆等小目标,分割结果中的轮廓保留更完整;对于植被区域,类间混淆有所减轻。这些现象说明,DFAM, DWFU 和 GSPM 的组合有助于改善复杂区域中的分割结果。

在 Potsdam 数据集上,本文方法 mIoU 达到 87.13%,虽较 RS3Mamba 低 0.14%,但参数量由 43.32 M 降低至 34.80 M,减少 8.52 M;其 GFLOPs 为 34.79,与 RS3Mamba 处于相近计算量级。受单卡 3090 资源限制,实验中统一采用批大小 8、早停 25 轮等一致设置,并严格遵循各对比方法官方参数与数据处理流程。尽管在极致精度上未突破,但通过级联精炼融合策略,在单一 Swin-T 轻量骨干下达到了接近双骨干 Mamba 模型的性能,体现了资源受限场景下的实用价值。

表 7 ISPRS Potsdam 数据集上不同分割网络实验结果对比

Tab. 7 Comparison of experimental results of different segmentation networks on the ISPRS Potsdam Dataset

网络	F1/%					mF1/%	OA/%	mIoU/%
	不透水地表	建筑	低矮植被	树	车			
PyramidMamba	93.06	96.45	87.37	88.42	96.01	92.42	91.25	86.22
GDGNet	93.58	96.84	88.29	89.47	96.53	92.94	91.56	87.06
UANet	93.31	96.76	88.40	89.68	96.28	92.89	91.50	86.91
CMTFNet	93.60	97.11	88.06	89.04	96.86	92.93	91.54	87.03
AFENet	93.72	97.16	88.01	89.13	96.67	92.94	91.56	87.03
RS3Mamba	93.52	97.03	88.32	89.67	96.87	93.08	91.58	87.27
CR-Net	93.61	96.85	88.33	89.56	96.72	93.01	91.57	87.13

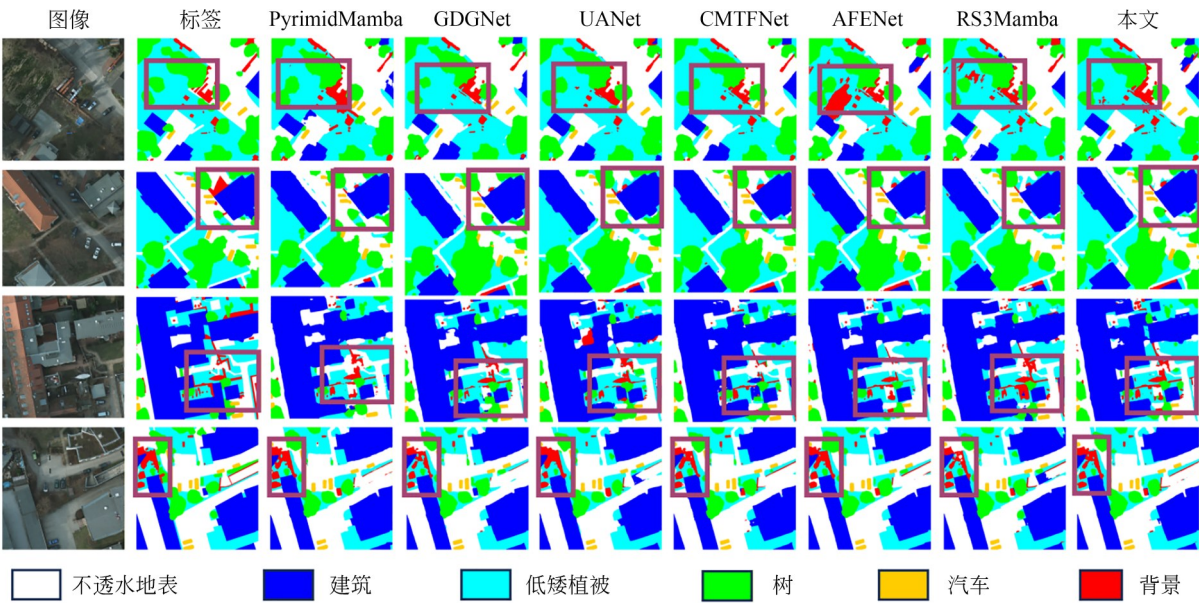


图 6 ISPRS Potsdam 数据集对比实验可视化

Fig. 6 ISPRS Potsdam Dataset comparison experiment visualization

3. 5. 2 在 ISPRS Vaihingen 数据集上的对比实验

为了验证 CR-Net 的泛化能力,本文在 ISPRS Vaihingen 数据集上与多种先进语义分割模型进行了对比。如表 8 所示,CR-Net 在该数据集上取得了 83.86% 的 mIoU,91.59% 的 OA 及 91.11% 的 mF1,三项整体指标均为所有对比模型中的最优,表明所提方法具有较好的跨数据集泛化能力。进一步从类别指标来看,CR-Net 在“低矮植被”类别上的 F1 分数达到 85.84%,为对比方法中的最高值;在“树”类别上取得

90.63% 的 F1 分数,仅低于 CMTFNet,表明模型对复杂植被纹理及相邻植被类别的区分仍存在提升空间;在“车”类别上取得 89.57% 的 F1 分数,仍保持了较强的竞争力,体现了模型在复杂场景下对小目标与结构性目标的较为稳定的分割能力。

图 7 的可视化结果与表 8 中的定量结果相对应。针对微小且离散的车辆目标(第 2 行红框所示,彩图见期刊电子版),CR-Net 相比 PyramidMamba 与 CMTFNet,其漏检与粘连现象有所减少,车辆轮廓表达更完整;在处理规则

表 8 ISPRS Vaihingen 数据集上不同分割网络实验结果对比

Tab. 8 Comparison of experimental results of different segmentation networks on the ISPRS vaihingen dataset

网络	F1/%					mF1/%	OA/%	mIoU/%
	不透水地表	建筑	低矮植被	树	车			
PyramidMamba	92.54	95.53	84.35	89.83	85.54	89.56	90.73	81.36
DFCGRNet	92.87	95.79	84.58	89.97	89.34	90.51	91.06	82.83
UANet	92.68	95.54	84.39	89.80	88.90	90.26	90.76	82.46
CMTFNet	91.78	95.97	84.71	90.96	89.40	90.39	90.12	82.66
AFENet	93.03	96.01	84.90	90.31	89.19	90.69	91.20	83.18
RS3Mamba	92.46	95.34	84.39	89.92	87.60	89.94	90.61	81.94
CR-Net	93.32	96.17	85.84	90.63	89.57	91.11	91.59	83.86

人造地物时(第 4 行),相比于 GDGNet 等模型

分割掩膜边界形态相对更平直,几何结构更

存在的边缘波浪化与侵蚀问题,CR-Net 生成的

完善。

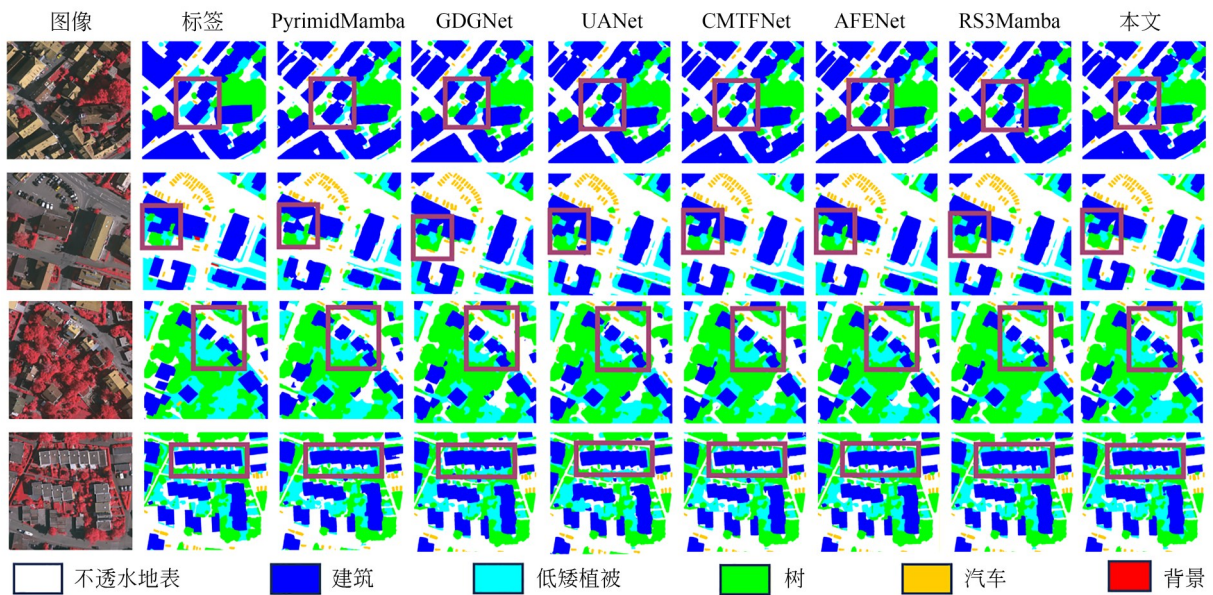


图7 ISPRS Vaihingen数据集可视化结果

Fig. 7 ISPRS Vaihingen Dataset comparison experiment visualization

3. 5. 3 在 LoveDA 数据集上的泛化实验

为了进一步考察模型在复杂场景中的表现,本文在 LoveDA 数据集上进行了泛化实验。此处选取的对比模型为在 Potsdam 和 Vaihingen 数据集上与本文模型性能接近的 RS3Mamba 和 AFENet。实验结果如表 9 所示,CR-Net 在该数

据集上取得了最优综合性能,mF1,OA 和 mIoU 分别达到 69.64%,70.79% 和 54.26%,均优于两种对比方法。其中,CR-Net 在建筑、道路、水面、森林和农田类别上的 F1 分别达到 79.60%,74.08%,81.92%,61.72% 和 67.89%,体现出较强的目标识别与边界刻画能力。

表 9 LoveDA 数据集泛化实验结果对比

Tab. 9 Generalization of experimental results of different segmentation networks on the LoveDA dataset

网络	F1/%							mF1/%	OA/%	mIoU/%
	背景	建筑	道路	水面	荒地	森林	农田			
AFENet	70.94	76.19	72.60	81.80	52.14	60.05	67.71	68.77	70.39	53.15
RS3Mamba	68.09	74.81	65.59	76.88	44.56	58.47	70.06	65.49	68.27	49.50
CR-Net	70.77	79.60	74.08	81.92	51.52	61.72	67.89	69.64	70.79	54.26

如图 8 所示,高亮区域进一步验证了 CR-Net 在复杂场景下的判别优势:第一行和第四行表明其对密集建筑的结构保持更完整,能够更准确地区分相邻建筑边界,减少粘连与破碎;第二行展现了道路区域更优的连续性和边缘规则性,降低了与周边背景或裸地的混淆;第三行则说明在森林

与农田交错区域中,CR-Net 对地物范围的刻画更完整,能够更有效缓解类别混淆。总体来看,CR-Net 在建筑、道路、水面等关键类别上优势明显;在森林类别上取得最高 F1,而农田类别仍存在一定提升空间;所展示样例表明,其在部分森林—农田交错区域能够形成较完整的区域预测。

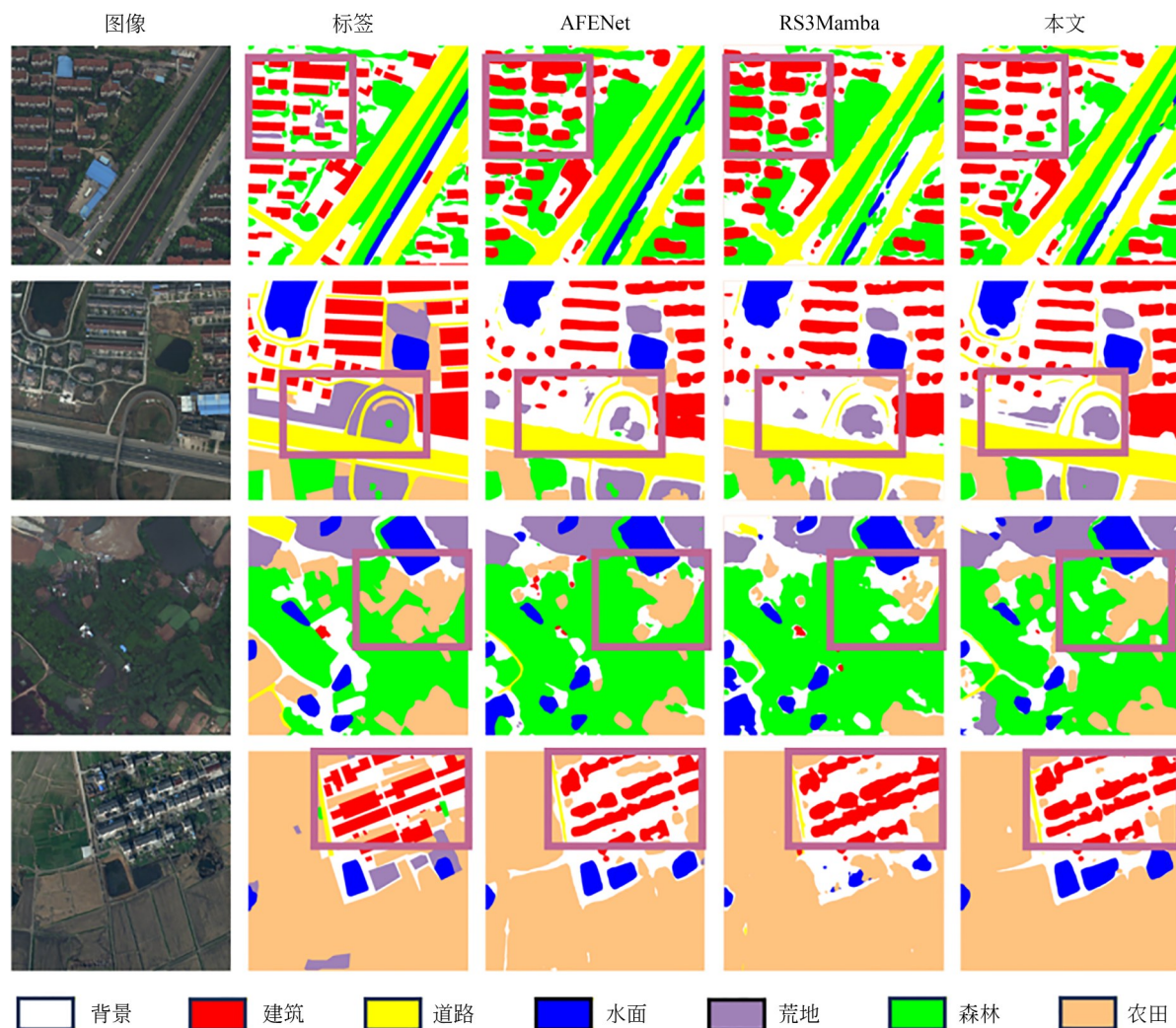


图 8 LoveDA 数据集泛化实验可视化结果

Fig. 8 LoveDA Dataset generalization experiment visualization

4 结 论

针对高分辨率遥感图像语义分割中细节信息丢失、边界模糊以及结构性目标易出现局部断裂和轮廓失真的问题,本文提出了一种基于 Swin-Transformer 编码器的级联精炼网络 CR-Net。该网络在解码阶段依次构建对偶特征聚合模块、动态小波融合上采样模块和几何结构感知模块,实现跨层语义互补、频域细节恢复与末端几何校准,从而提升目标边界和结构信息的表达能力。

在 ISPRS Potsdam, ISPRS Vaihingen 和 LoveDA 数据集上的实验结果表明,CR-Net 的 mIoU 分别达到 87.13%, 83.86% 和 54.26%。消融实验进一步验证了 DFAM, DWFU 和

GSPM 的有效性,三者联合使用时较基线模型取得更优性能,说明级联精炼策略能够有效改善复杂建筑轮廓、小目标保持及跨场景分割效果。

尽管如此,CR-Net 在极端复杂场景下仍存在一定局限。当影像中存在密集建筑、阴影遮挡、尺度剧烈变化及光谱纹理相近地物时,模型仍可能出现局部误分、边界粘连或小目标漏检。LoveDA 数据集中荒地、农田等弱结构类别的 F1 分数相对较低,说明模型对复杂域差异和类别间模糊边界的适应性仍需增强。此外, Swin-Transformer 编码器及多级精炼模块增加了一定计算开销,端侧实时部署仍有优化空间。未来可从困难类别建模、边界一区域协同约束、模型轻量化、知识蒸馏及域自适应等方面进一步改进,以提升

模型的泛化能力和工程应用价值。

作者贡献声明:

刘春娟:论文审核与编辑;

赵浩然:论文构思和写作;

闫浩文:提供资源;

吴小所:获取资助和指导;

包易航:辅助可视化。

参考文献:

- [1] 李智杰, 宋易宸, 李昌华, 等. 面向遥感图像超分辨率重建的跨尺度余弦注意力网络[J]. 计算机工程与应用, 2026, 62(1): 285-296.
LI Z J, SONG Y C, LI C H, *et al.* Cross-scale cosine attention network for remote sensing image super-resolution reconstruction [J]. *Computer Engineering and Applications*, 2026, 62(1): 285-296. (in Chinese)
- [2] 周冬梅, 贾琴琴. 融合 Mamba 的多尺度遥感图像实例分割网络[J]. 光学精密工程, 2025, 33(24): 3940-3955.
ZHOU D M, JIA Q Q. Multi-scale remote sensing image instance segmentation network with mamba integration [J]. *Opt. Precision Eng.*, 2025, 33(24): 3940-3955. (in Chinese)
- [3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 7-12, 2015, Boston, MA, USA. IEEE, 2015: 3431-3440.
- [4] CHEN L C, PAPANDREOU G, SCHROFF F, *et al.* Rethinking Atrous Convolution for Semantic Image Segmentation [EB/OL]. 2017: *arXiv*: 1706.05587. <https://arxiv.org/abs/1706.05587>
- [5] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Cham: Springer, 2015: 234-241.
- [6] HU J, SHEN L, SUN G. Squeeze-and-excitation Networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [7] YAN W D, CAO L, YAN P, *et al.* Remote sensing image change detection based on swin transformer and cross-attention mechanism[J]. *Earth Science Informatics*, 2024, 18(1): 106.
- [8] TENG J, ZHU Z W, PU D M, *et al.* MSCSC-Net: multiscale convolution selective context network with boundary supervision for semantic segmentation of remote sensing images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 25892-25908.
- [9] WANG L B, LI R, ZHANG C, *et al.* UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 190: 196-214.
- [10] LIU X L, LIN S F, XU N N. FMDformer: frequency mixed decomposed transformer for time series forecasting [J]. *Concurrency and Computation: Practice and Experience*, 2025, 37(27/28): e70434.
- [11] WEI G Y, XU J D, YAN W Q, *et al.* Dual-domain fusion network based on wavelet frequency decomposition and fuzzy spatial constraint for remote sensing image segmentation [J]. *Remote Sensing*, 2024, 16(19): 3594.
- [12] FINDER S E, AMOYAL R, TREISTER E, *et al.* Wavelet Convolutions for Large Receptive Fields [M]. *Computer Vision-ECCV 2024*. Cham: Springer Nature Switzerland, 2024: 363-380.
- [13] SHANG X K, LI G H, JIANG Z Y, *et al.* Holistic Dynamic Frequency Transformer for image fusion and exposure correction [J]. *Information Fusion*, 2024, 102: 102073.
- [14] WU H L, HUANG P, ZHANG M, *et al.* CMT-FNet: CNN and multiscale transformer fusion network for remote-sensing image semantic segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 2004612.
- [15] LI X H, XIE L L, WANG C F, *et al.* Boundary-enhanced dual-stream network for semantic segmentation of high-resolution remote sensing images [J]. *GIScience & Remote Sensing*, 2024, 61(1): 2356355.
- [16] SONG Z X, LIU X, ZHANG W H, *et al.* SPG-

- Net: a shape-prior guided network for medical image segmentation [C]. *International Joint Conference on Artificial Intelligence.*, 2024
- [17] QI Y L, HE Y T, QI X M, *et al.* Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation [C]. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 1-6, 2023, Paris, France. IEEE, 2024: 6047-6056.
- [18] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674-693.
- [19] DAUBECHIES I. *Ten Lectures on Wavelets* [M]. Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 1992.
- [20] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [C]. *International Conference on Learning Representations.*, 2017
- [21] GAO F, FU M, CAO J C, *et al.* Adaptive frequency enhancement network for remote sensing image semantic segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5619415.
- [22] WANG K, ZHANG X B, WANG X M, *et al.* Gradient decoupling guided network for high-resolution remote sensing segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5623518.
- [23] LI J P, HE W, CAO W N, *et al.* UANet: an uncertainty-aware network for building extraction from remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5608513.
- [24] WANG L B, LI D X, DONG S J, *et al.* PyramidMamba: Rethinking pyramid feature fusion with selective space state model for semantic segmentation of remote sensing imagery [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2025, 144: 104884.
- [25] MA X P, ZHANG X K, PUN M O. RS3Mamba: visual state space model for remote sensing image semantic segmentation [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 6011405.

作者简介:



刘春娟(1970—),女,甘肃兰州人,教授,1996年于兰州交通大学获得学士学位,2004年于兰州交通大学获得硕士学位,现为兰州交通大学教授,主要从事语义分割与变化检测技术应用方向的研究。E-mail: liuchj@mail.lzjtu.cn

通讯作者:



吴小所(1981—),男,河南驻马店人,博士后,副教授。2004年于兰州交通大学获得学士学位,2011年于兰州交通大学获得硕士学位,2017年于兰州交通大学获得博士学位,现为兰州交通大学副教授,主要从事语义分割与变化检测技术应用方向的研究。E-mail: wuxs_laser@lzjtu.edu.cn