

引用格式: 梁书绮, 王雷, 孙燕青, 等. 基于自适应频率感知网络的遥感图像分割方法[J]. 电子科技大学学报, 2026, 55(1): 149-160.

LIANG S Q, WANG L, SUN Y Q, et al. Remote sensing image segmentation method based on adaptive frequency-aware network[J]. Journal of University of Electronic Science and Technology of China, 2026, 55(1): 149-160.

基于自适应频率感知网络的遥感图像分割方法



梁书绮¹, 王雷^{1*}, 孙燕青¹, 刘世龙¹, 杨善良¹, 李彬²

(1. 山东理工大学计算机科学与技术学院, 淄博 255000; 2. 华南理工大学自动化科学与工程学院, 广州 510641)

摘要: 现有遥感图像分割方法在跨尺度特征融合时缺乏对低频结构与高频细节的协同处理, 且无法对图像内容自适应响应, 导致难以有效解决因自然变化、光照和阴影干扰引起的高分辨率遥感图像中类内差异大、类间差异小的问题。为此提出了一种自适应频率感知网络 (AFANet)。首先, 提出一种频率动态融合模块, 通过自适应低通和高通滤波在保留低频结构的同时抑制高频噪声分量, 并增强高频细节边界信息。其次, 构建双域学习模块集成空间和频率信息, 实现空间域局部细节与频域全局结构的联合建模。最后, 引入一个细节增强模块, 利用不同的差分卷积以增强模型的特征提取和泛化能力。在 Vaihingen 和 Potsdam 两个经典公开数据集上通过对比和消融实验的定量及可视化分析表明, AFANet 在 F1 分数、总体精度和平均交并比等指标中优于 7 种先进的分割方法, 验证了 AFANet 的优越性能。

关键词: 遥感图像; 语义分割; 特征融合; 自适应频率滤波; 空间域—频域 Transformer
中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.12178/1001-0548.2025085

Remote sensing image segmentation method based on adaptive frequency-aware network

LIANG Shuqi¹, WANG Lei^{1*}, SUN Yanqing¹, LIU Shilong¹, YANG Shanliang¹, and LI Bin²

(1. School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China;

2. School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: The existing remote sensing image segmentation methods lack the coordinated processing of low-frequency structures and high-frequency details during cross-scale feature fusion, and cannot adaptively respond to image content, making it difficult to effectively solve the problem of large intra-class variations and small inter-class variations in high-resolution remote sensing images caused by natural changes, illumination, and shadow interference. To address these problems, an adaptive frequency-aware network (AFANet) is proposed and implemented. First, a frequency dynamic fusion module is proposed, which suppresses high-frequency noise components while retaining low-frequency structures through adaptive low-pass and high-pass filter, and enhancing high-frequency detail boundary information. Secondly, a dual-domain learning block is constructed to integrate spatial and frequency information to achieve joint modeling of local details in the spatial domain and global structures in the frequency domain. Finally, a detail-enhanced module is introduced to enhance the feature extraction and generalization capabilities of the model using different differential convolutions. The quantitative analysis and visualization results of the comparative and ablation experiments on the two classic public datasets, Vaihingen and Potsdam, show that AFANet outperforms seven state-of-the-art segmentation methods in terms of F1 score, OA and mIoU, with the mIoU reaching 85.13% and 87.81% respectively, verifying the superior performance of AFANet.

Key words: remote sensing images; semantic segmentation; feature fusion; adaptive frequency filter; spatial-frequency Transformer

高分辨率遥感图像提供了详尽的地表特征信息, 已成为农业监测^[1]、地理学研究^[2]、城市规划^[3]

及军事侦察^[4]等领域不可或缺的信息来源。遥感图像语义分割旨在将图像中的每个像素分配至特定的

收稿日期: 2025-04-27

基金项目: 国家自然科学基金面上项目 (62273155); 山东省自然科学基金 (ZR2021MF017, ZR2024QF022); 山东省重点研发计划 (2023RKY01015)

作者简介: 梁书绮, 主要从事计算机视觉方面的研究。

*通信作者 E-mail: wanglei0511@sdut.edu.cn

地理空间对象类别, 从而实现地对地表特征的精准识别与提取。与自然图像相比, 遥感图像语义分割面临多重挑战。首先, 遥感图像因不同种类间颜色、形状相似度较高, 导致类间差异较小, 模型难以区分相似的类别。其次, 遥感图像因特有的光照条件, 许多不规则的阴影会导致同一类别中的颜色变化, 且复杂的空间结构会导致不同类别间的相互遮挡, 导致提取的语义特征存在较大偏差, 较大的类内方差不利于语义的识别。因此, 如何解决这些问题已成为研究热点。

传统的遥感图像分割方法, 如阈值分割^[5]、边缘检测^[6]和聚类^[7]等经典算法, 主要依赖于像素的颜色、纹理等属性。传统遥感图像分割方法虽在特定条件下表现良好, 但面对复杂多变的遥感图像时存在局限性。因此, 深度学习逐渐在遥感图像分割领域得到应用^[8]。UNet 网络凭借其独特的编码器和解码器架构成为分割领域中的主要模型框架^[9]。文献 [10] 引入残差连接、亚属性卷积和金字塔场景解析池, 设计的 ResUNet 在高分辨率航空图像分割上表现良好。CNN 网络对于图像局部信息的学习有优势, 但缺乏上下文交互能力, 全局信息获取明显不足。近年来, Transformer 在计算机视觉领域展现出显著优势且对于图像分割十分有效。Vision Transformer (ViT)^[11]、Swin Transformer^[12]、SETR^[13] 等利用自注意力机制来获取全局上下文信息, 提高了语义分割的性能。文献 [14] 提出一种长短距离 Transformer, 可以在模型的所有阶段学习局部和全局表示。得益于 CNN 和 Transformer 优势的互补, 当前 CNN-Transformer 混合架构的模型在分割任务中表现更为出色。UNetFormer^[15] 是一种针对城市场景构建的分割模型, 在 Potsdam 和 Vaihingen 数据集上表现出优秀的分割结果。CMTFNet^[16] 将 CNN 和多尺度 Transformer 融合用于提取高分辨率遥感图像的局部信息和多尺度全局上下文信息。

尽管现有 CNN-Transformer 框架模型通过局部卷积与全局注意力机制有效融合了空间域的多尺度特征, 但对于频域信息的提取与建模仍存在局限性。在频域中, 高频分量对应遥感图像的快速变化区域, 高频噪声 (如光照引起的反光、阴影遮挡) 造成特征提取过程中过度聚焦局部变异而忽视区域一致性, 容易导致模型将单一连续区域碎片化, 进而引发类内预测不连续问题。低频分量则对应遥感图像中缓慢变化的区域 (如大面积植被的平滑轮廓、

水体的均匀纹理)。如果对低频分量的表征不足, 削弱了对整体结构的感知, 模型难以通过衰减的低频响应准确边界定位。同时, 边界轮廓和锐利纹理等重要高频分量在下采样过程中容易被大量背景噪声淹没, 最终引发分割边缘的模糊与偏移。因此, 将频域信息有效整合到空间分割网络中至关重要。

频域分析在信号处理中被广泛应用, 深度学习图像处理已经扩展至遥感图像分割领域。文献 [17] 提出了一个空间和频率双域变换器, 引入了快速傅里叶变换, 将图像变换到频域进行特征提取, 弥补空间域的不足。文献 [18] 整合了可变形卷积和傅里叶变换用于遥感道路图像的分割。文献 [19] 根据离线余弦变换提取的频率信息设计了频域注意力机制, 用于遥感图像分割。文献 [20] 利用小波变换提取低频分量、水平高频分量、垂直高频分量和对角线高频分量, 结合空间和频域特征提升了遥感图像分割效果。文献 [21] 运用一种动态频率滤波块, 采用不同卷积大小的深度卷积层组成高通滤波器, 并利用平均池化作为低通滤波器。该方法可以在空间域捕获重要的高频和低频特征, 文献 [22] 将这种方法运用在遥感图像分割领域, 并获得了较好的结果。

基于频域信息的遥感图像分割方法已被证明具有巨大潜力, 但大都基于固定的频域分析方式提取频率信息, 由于遥感图像的复杂性, 不同区域的低频成分存在显著差异, 难以动态适配图像内容特征。同时, 现有频域分析方式普遍存在于编码器与解码器中, 在跨尺度特征融合阶段未整合高低频信息, 导致分割结果在边界精度与类内一致性上表现受限。因此, 本文提出并实现了一种自适应频率感知网络 (adaptive frequency-aware network, AFANet)。在编码器端采用 ConvNeXt 作为主干网络并在解码器构建双域学习模块 (dual-domain learning block, DDLB), 在频域收集全局上下文信息, 在空间域学习图像的局部特征, 实现跨域特征的双向交互与互补增强。用频率动态融合模块 (frequency dynamic fusion module, FDFM) 在跳跃连接处进行特征融合。通过局部自适应调整滤波权重, 降低全局噪声和无关结构的干扰。对低分辨率解码器特征图实施自适应低通滤波, 在保留低频结构的同时抑制高频噪声分量并在上采样时平滑特征, 降低类内不一致性。对高分辨率编码器特征图实施自适应高通滤波, 提取边界轮廓等重要高频细节, 增强类间差异性。本文引入一个细节增强模块 (detail-enhanced

module, DEM), 利用差分卷积增强模型的特征提取和泛化能力, 减轻光照阴影的干扰。

1 AFANet 模型

1.1 总体架构

AFANet 网络的总体架构如图 1 所示, 整体采用 UNet 编码器和解码器框架。在编码器端, 采用 ConvNeXt^[23] 作为主干网络。该网络基于 ResNet 的结构基础, 通过引入分层转换、逆瓶颈层和更深

的网络结构优化, 实现更高效的特征提取能力。为了减少参数, 使用 Tiny 版本的 ConvNeXt 进行模型设计。输入图像经 4 级下采样后生成分辨率逐级递减的 4 组特征图 $E_i, i \in \{1, 2, 3, 4\}$ 。对后两级特征图 E_3 和 E_4 分别提取高频信息和低频信息进行跨尺度特征融合, 将特征图送入解码器中。在解码器端, 每个双域学习模块与上一级的编码器 ConvBlock 模块再进行多尺度特征融合, 最后一个双域学习模块的特征图经过细节增强模块后得到最终的结果。

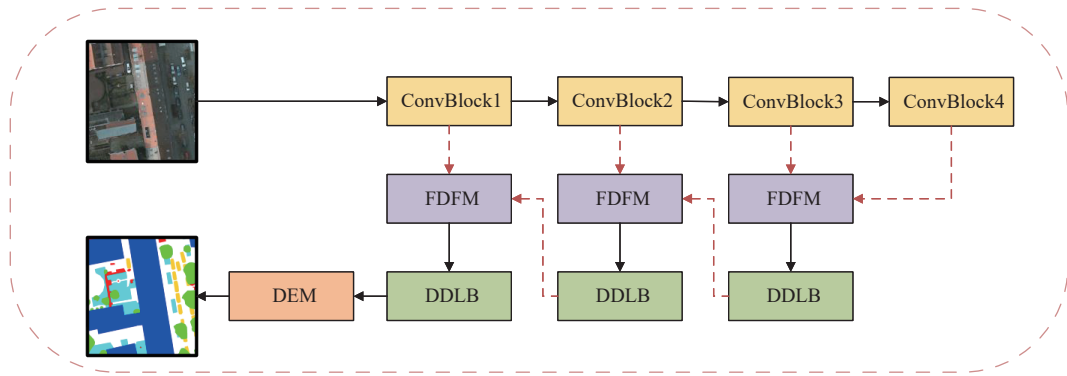


图 1 AFANet 网络总体框架

1.2 频率动态融合模块

在 UNet 框架中, 跳跃连接通过将解码器输出的低分辨率特征图与编码器对应层级的高分辨率特征图进行跨尺度特征融合, 有效缓解了深层语义信息与浅层空间细节间的特征鸿沟。转置卷积、双线性插值和最近邻插值等都是常见的跳跃连接上采样方法来实现分辨率对齐, 但由于它们都是固定上采样策略并非学习型的, 上采样核权重与输入内容无关, 难以适配遥感图像复杂场景下的多尺度特征分布。并且, 简单的插值上采样可能会传播局部特征不一致性, 扩大类内差异。受文献 [24-26] 的启发, 频率动态融合模块采用动态生成上采样核来实现特征图的上采样过程, 根据输入特征的内容自适应调整上采样策略。在此基础上, 为了收集高低频信息, 基于特征图内容自适应生成低通与高通滤波器, 并为每个坐标预测周边空间变化的滤波参数, 通过局部自适应调整滤波权重, 有效抑制全局噪声和无关结构的干扰。从低分辨率高级特征图中提取低频结构信息, 在保留低频结构的同时抑制高频噪声分量; 从高分辨率低级特征图中提取重要高频细节, 以恢复在下采样和上采样期间因分辨率变化而减少的高频边界细节, 提升边界清晰度和纹理表现。通过这种自适应的频率分解与融合机制, 有效桥接了低级特征与高级特征, 显著提升了跨尺度特征融合的代表能力。

如图 2 所示, 频率动态融合模块首先对特征图进行通道压缩, 将原始 C 维特征投影至低维空间 C_m 。此操作在维持模型表达能力的前提下, 降低了后续滤波参数生成的计算负荷, 释放了计算资源。针对解码器低分辨率的特征图根据每个像素点的空间位置生成重组核:

$$M = \psi(F_{de}^{comp}), M \in \mathbb{R}^{k^2 \times \sigma H \times \sigma W} \quad (1)$$

式中, F_{de}^{comp} 表示经过通道压缩后的解码器特征图; k 是内核大小; σ 是特征图上采样倍数。随后通过 Softmax 函数与归一化约束, 确保滤波器权重非负性且总和为 1, 从而将重组核 M 转换成具有平滑特性的低通滤波核 M_{LP} 。最后进行特征重组, 对于目标输出位置 (i, j) , 在输入特征图上提取对应的局部邻域 $N(i, j)$, 邻域大小与核尺寸 $k \times k$ 一致, 在低通滤波核中 k 值设置为 3。根据低通滤波核 M_{LP} 对局部邻域进行加权求和:

$$F_{low}^{(i,j)} = \sum_{(i,j) \in N(i,j)} M_{LP} \cdot F_{de}^{(i+p,j+q)} \quad (2)$$

在邻域 $N(i, j)$ 内, 通过动态权重分配实现像素级自适应聚合, 重构后的特征图通过频域约束机制有效地提取到了低频信息并抑制远距离响应, 强化了语义表征一致性。利用高分辨率的编码器特征图提取高频特征的步骤与之前一致, 采用 5×5 大小的重构核得到低频特征后, 利用减法增强高频细节:

$$F_{en}^{high} = F_{en} + (F_{en} - F_{en}^{low}) \quad (3)$$

最终融合特征结果通过增强后的高频特征与上

采样的低频特征相加得到:

$$F_{out} = F_{de}^{low} + F_{en}^{high} \quad (4)$$

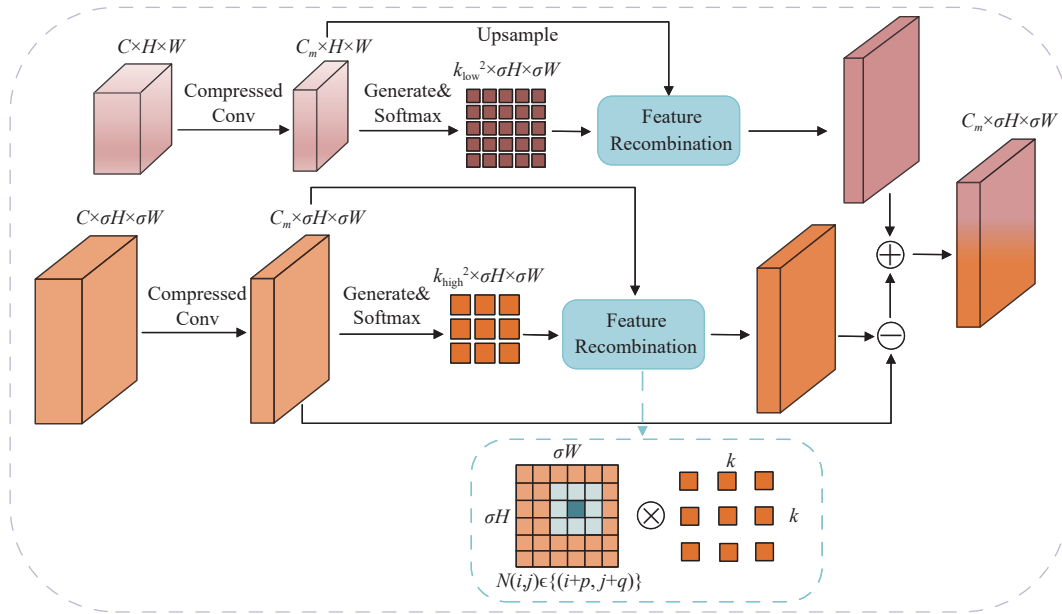


图 2 频率动态融合模块结构图

1.3 双域学习模块

将频率动态融合模块输出的特征图送入双域学习模块。双域学习模块通过全局分支的双频注意力模块和局部分支的通道空间协同模块对特征图进行学习。通过频域全局建模与空间域局部增强的协同设计,实现全局-局部特征互补与多域信息融合,具体结构如图 3 所示。

1) 双频注意力模块 (bi-frequency attention module, BFAM)。传统的方法中高频与低频信息常在同一特征中耦合,导致自注意力机制难以区分特征图的具体细节与整体结构。双频注意力模块将高频与低频信息解耦,这样既能双分支独立优化不同频段特征,提升对多尺度目标的捕捉能力,经相加融合后又可以形成互补,避免单分支的过平滑或过锐化问题。

双频注意力模块采用与频率动态融合模块相同的方式分别使用自适应高通滤波与低通滤波提取高频特征和低频特征。利用提取出的高频特征作为查询向量 (Query),而原始特征图充当键 (Key) 和值 (Value),这样可以使得模型专注于学习和增强图像中的细节部分。类似地,使用低频特征作为查询向量,原始特征图同样用作键和值。这样有助于模型更好地理解图像的整体结构和背景信息。在特征映射过程中,Key 和 Value 只需使用一组共享权重以降低参数量,公式如下:

$$\begin{aligned} Q_i &= W_i^Q \cdot F_i \quad K = W_K \cdot F \\ V &= W_V \cdot F \quad i \in \{\text{low, high}\} \end{aligned} \quad (5)$$

高频分支权重 W_{high}^Q 倾向于学习高频敏感的投影方向,低频分支权重 W_{low}^Q 则聚焦低频特征映射。随后分别进行多头注意力的计算,相加后得到在高低频影响下的新特征图:

$$F_{out} = \sum_{i \in \{\text{low, high}\}} \text{softmax} \left(\frac{Q_i \cdot K^T}{\sqrt{d_k}} \right) \cdot V \quad (6)$$

尽管以上操作能有效建模高频细节与低频结构的关联性,但其受限于窗口划分机制,难以捕获跨窗口的长程空间依赖。因此引入多尺度条形卷积,通过并行水平、垂直方向的长条形卷积核,补偿窗口自注意力机制的局部性局限,促进窗口之间的交互。不同尺度的卷积核形成多级感受野,协同增强全局上下文感知。小型的条形卷积聚焦局部邻域,捕捉细粒度方向特征 (如短道路段、汽车);大型的覆盖更大范围,建模跨区域的远距离关联 (如大片农田、大型建筑)。

2) 通道空间协同模块 (channel-spatial synergy module, CSSM)。为了提取遥感图像的局部特征,提出了一种新颖的通道空间协同模块。对通道和空间两种注意力进行串行操作,在通道注意力操作中,为了克服传统全局池化方法导致的信息丢失问题,空间通道协同模块沿宽度和高度两个方向分

别对输入张量进行平均池化操作, 并采用一维深度可分离卷积进行处理: 先进行深度卷积 (depthwise conv), 使用大小为 5 的一维卷积核沿长和宽维度进行特征提取; 再进行逐点卷积 (pointwise conv) 融合通道信息。经过 Sigmoid 函数后生成通道注意力权重。利用一维深度可分离卷积来提取通道特征, 而不是采用降维和增维的操作, 从而确保了通道特征的完整性和信息的精确性。这样有助于更好地捕捉通道之间的依赖关系, 提升模型的表现力。随后, 将通道注意力输出的特征图进行空间注意力

操作。在空间注意力阶段, 首先采用多尺度卷积策略, 使用 3 个大小分别为 3×3 、 5×5 和 7×7 的并行卷积分别捕捉不同尺度下的重要空间信息, 再进行特征融合。接着, 应用实例归一化 (instance normalization, IN) 对融合后的特征图进行归一化处理, 同样经过 Sigmoid 函数生成单通道空间注意力权重, 针对单个图像的每个通道独立地进行归一化, 使得模型更加关注于图像内部的局部特征而非全局信息。最后, 将空间注意力权重与原始特征图相乘, 得到最终输出的特征图。

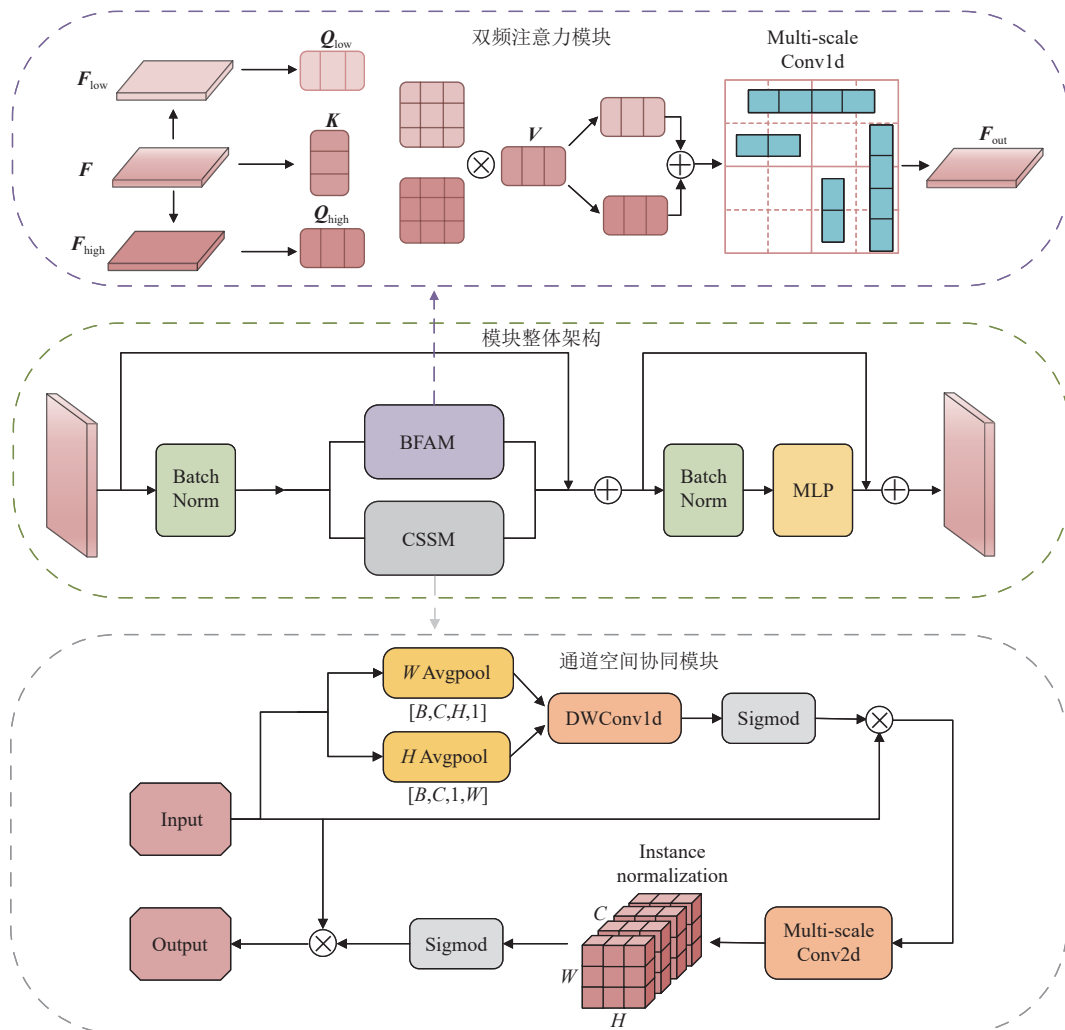


图 3 双域学习模块结构图

1.4 细节增强模块

在编码器首层与解码器末层的特征融合后, 引用一个细节增强模块 (detail-enhanced module, DEM) 进一步减轻由于光照导致的阴影问题。该模块包含 1 个普通卷积和 3 个差分卷积, 通过差分卷积来增强模型对图像局部变化的敏感度。差分卷积最先是用于边缘检测^[27]、人脸防伪任务^[28] 以及单图像去雾^[29] 等领域。鉴于遥感图像特有的挑战——由于

复杂的光照条件导致的阴影问题, 使得物体边界模糊不清, 这直接影响了图像分割的准确性。因此, 差分卷积所具备的能力对于解决此类问题尤为关键。

3 个差分卷积分别采用中心差分卷积 (central difference convolution, CDC)、水平差分卷积 (horizontal difference convolution, HDC) 和垂直差分卷积 (vertical difference convolution, VDC)。中心差分卷积通过计算中心像素与周围像素之间的差

异来增强对局部变化的敏感度。卷积核权重矩阵为:

$$\mathbf{W}_{CD} = \begin{bmatrix} w_0 & w_1 & w_2 \\ w_3 & w_4 - \sum_{k=0}^8 w_k & w_5 \\ w_6 & w_7 & w_8 \end{bmatrix} \quad (7)$$

水平差分卷积提取水平方向梯度, 构建左右对称的权重抵消关系, 以此检测水平方向的变化。其卷积核权重矩阵为:

$$\mathbf{W}_{HD} = \begin{bmatrix} w_0 & 0 & -w_0 \\ w_3 & 0 & -w_3 \\ w_6 & 0 & -w_6 \end{bmatrix} \quad (8)$$

垂直差分卷积通过显式建模垂直方向的特征梯度变化, 构建上下行权重抵消关系。其卷积核权重矩阵为:

$$\mathbf{W}_{VD} = \begin{bmatrix} w_0 & w_1 & w_2 \\ 0 & 0 & 0 \\ -w_0 & -w_1 & -w_2 \end{bmatrix} \quad (9)$$

利用卷积层的可加性来简化并行部署的卷积, 将这些卷积核相加获得一个相同输出的等效核, 最终特征图只需要进行一次卷积运算即可, 公式为:

$$\mathbf{F}_{out} = \sum_{i=1}^4 \mathbf{F} * \mathbf{W}_i = \mathbf{F} * \sum_{i=1}^4 \mathbf{W}_i = \text{Conv}(\mathbf{F}) \quad (10)$$

式中, * 代表卷积操作。

1.5 损失函数

为优化 AFANet 的训练过程, 采用复合损失函数策略, 它包含一个主损失函数与一个辅助损失函数。其中, 主损失函数采用 Dice 损失与交叉熵损失结合的方式, 有效提升小目标分割精度的同时避免模型偏向大面积的背景预测。Dice 损失与交叉熵损失函数分别如下:

$$\mathcal{L}_{dice} = 1 - \frac{2}{N} \sum_{n=1}^N \sum_{k=1}^K \frac{\widehat{y}_k^{(n)} y_k^{(n)}}{\widehat{y}_k^{(n)} + y_k^{(n)}} \quad (11)$$

$$\mathcal{L}_{ce} = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^{(n)} \log \widehat{y}_k^{(n)} \quad (12)$$

$$\mathcal{L}_{main} = \mathcal{L}_{dice} + \mathcal{L}_{ce} \quad (13)$$

式中, N 和 K 分别代表数据集的样本数和类别数; $y^{(n)}$ 表示真实标签的 one-hot 编码, 其对应的网络输出为 softmax 层生成的概率分布 $\widehat{y}^{(n)}$ 。辅助函数针对频率动态融合模块, 将 3 个频率动态融合模块的结果经过上采样后相加, 选择交叉熵损失作为辅助函数 \mathcal{L}_{aux} 并乘以系数因子 α , α 的值设置为 0.4。最终完整的损失函数公式为:

$$\mathcal{L} = \mathcal{L}_{main} + \alpha \mathcal{L}_{aux} \quad (14)$$

2 实验与结果

2.1 数据集与实验设置

采用国际摄影测量与遥感学会 ISPRS 提供的两个数据集 Potsdam 和 Vaihingen 对 AFANet 网络进行实验。两个数据集都被分类为 6 种最常见的土地覆盖类型, 包括不透水表面、建筑物、低矮植被、树木、汽车和杂波 (背景, 包括水体和其他不感兴趣的语义对象)。

Potsdam 聚焦于德国波茨坦地区, 是一个典型的都市遥感数据集。图像中的物体大小差异较大, 有一定的分割难度, 共有 38 幅航空遥感影像, 分辨率为 $6\,000 \times 6\,000$ 像素, 地面采样距离 (GSD) 为 5 cm, 每幅影像由正射影像 (TOP) 以及数字表面模型 (DSM) 和归一化数字表面模型 (NDSM) 组成, 其中正射影像具有 3 个多光谱波段 (近红外, 红色, 绿色)。选用正射影像进行模型训练并在使用前将图像裁剪成 $1\,024 \times 1\,024$ 大小, 其中选定 23 幅图像用于训练, ID 分别为 2_10、2_11、2_12、3_10、3_11、3_12、4_10、4_11、4_12、5_10、5_11、5_12、6_7、6_8、6_9、6_10、6_11、6_12、7_7、7_8、7_9、7_11 和 7_12。剩余图像用于测试。

Vaihingen 数据集包含一个相对较小的村庄区域, 具有许多独立的建筑和小多层建筑, 同时也包含一些自然景观元素, 由 33 个平均大小为 $2\,494 \times 2\,064$ 像素的航空遥感影像组成, 地面采样距离 (GSD) 为 9 cm。该数据集与 Potsdam 数据集组成相同, 同样选用正射影像 (TOP) 进行模型训练并将图像裁剪成 $1\,024 \times 1\,024$ 大小。选取 ID 为 1、3、5、7、11、13、15、17、21、23、26、28、30、32、34、37 的 16 幅图像用于训练, 剩余的用于测试。

在 Ubuntu 操作系统下, 使用 PyTorch 框架在单张 NVIDIA GTX 3090 显卡上进行实验。使用 AdamW 优化器来训练实验中的所有模型。学习率设置为 6×10^{-4} 。采用余弦策略调整学习速率, 对训练数据采用随机缩放、随机翻转、随机旋转等数据增强技术。

为了评估 AFANet 网络的分割性能, 采用 F1 分数、总体精度 (over accuracy, OA)、平均交并比 (mean intersection over union, mIoU) 这 3 个评价指标。每个像素的预测结果分为真阳性 (TP)、假阳性 (FP)、真阴性 (TN) 和假阴性 (FN),

具体公式如下:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (15)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (16)$$

$$F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

$$OA = \frac{1}{M} \sum_{i=1}^M \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i} \quad (18)$$

$$mIoU = \frac{1}{M} \sum_{i=1}^M \frac{TP_i}{TP_i + FP_i + FN_i} \quad (19)$$

2.2 实验与讨论

如表 1 所示, 在 Potsdam 数据集上, AFANet 在所有指标上均表现最佳。具体而言, 与 MIFNet 相比, AFANet 的 F1 分数提高了 0.58%, 总体精度 (OA) 提高了 0.55%, 平均交并比 (mIoU) 提高

了 0.51%。图 4 选取了测试数据集中的两张局部放大图进行可视化对比。在第 1 张图像中, 卡车的货箱因颜色与道路相似而容易被误分为道路。FDNet 仅使用了傅里叶卷积, 并没有单独分离高频与低频信息, 因此对于卡车的轮廓与整体结构识别较差。Xnet 使用的小波变换可以进一步将二者区分, UperNet-FDAC 与 MIFNet 使用的平均池化代替低频能够将卡车边缘划分, 但结果均不尽人意。AFANet 利用自适应低频滤波根据图像局部区域的特性动态调整权重, 因此受旁边道路的影响更小。同理卡车旁边的汽车受到树木的遮挡, AFANet 仍能通过车身大致形状等低频信息推断出整体结构, 避免因局部遮挡导致无法识别。在第 2 张图像中, 固定的小波变换方式与简单的平均池化很难将形状与颜色极为相近的低矮植被与树木进行分辨, AFANet 保留树木的低频结构, 通过动态权重抑制低矮植被大片平滑区域的低频干扰, 因此分割效果更好。

表 1 Potsdam 数据集上的对比实验结果

Method	Per-class F1					MeanF1	OA	mIoU
	Imp. Sur.	Building	Low Veg.	Tree	Car			
UNet ^[9]	90.73	95.35	85.05	85.97	91.49	89.72	88.36	81.57
PSPNet ^[30]	91.35	95.17	84.36	83.48	90.34	88.94	88.98	80.63
UNetFormer ^[15]	93.60	97.20	87.70	88.90	96.50	92.80	91.30	86.80
FDNet ^[31]	93.51	96.76	87.20	88.78	96.13	92.48	91.34	86.58
Xnet ^[32]	91.98	96.80	86.93	88.76	95.81	92.05	90.56	85.51
UperNet-FDAC ^[33]	93.05	96.50	87.33	88.62	95.67	92.23	91.28	86.89
MIFNet ^[22]	93.63	97.07	87.78	89.42	96.20	92.80	91.80	87.30
AFANet(本文)	94.67	97.22	88.45	89.71	96.88	93.38	92.35	87.81

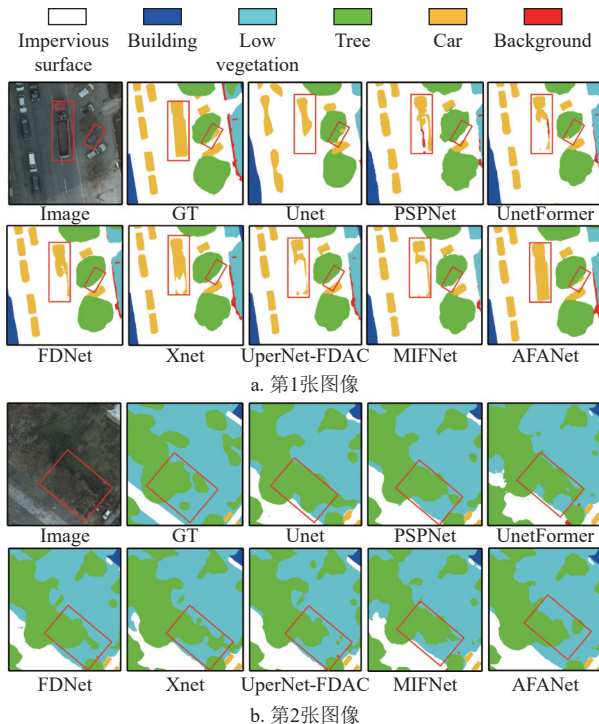


图 4 Potsdam 数据集上的可视化对比图

表 2 列出了 Vaihingen 数据集上每个网络的评估指标, 其中 AFANet 是表现最佳的网络。具体而言, 与最新的 MIFNet 网络相比, AFANet 的 F1 分数提高了 0.35%, 总体精度 (OA) 提高了 1.25%, 平均交并比 (mIoU) 提高了 0.6%。此外, 还显示了预测结果中每个类别的 F1 分数, 涵盖 6 个分类, 其中把杂乱类别 (背景) 排除掉。通过局部放大可视化图 5 发现, 受益于从图像中提取到的高频信息与改进的通道空间注意力机制, AFANet 对于低矮灌木和树木与其他类别在边界处的分割效果比其他网络更优, 有助于模型精准定位复杂边界。在光照阴影影响下, AFANet 使用的差分卷积可以通过计算像素间的差异, 强化模型对局部细微变化的敏感度, 对于面积较小的汽车, 哪怕阴影将车辆全部覆盖, AFANet 也能做到更准确的识别。通过引入的多尺度条形卷积, 在关注局部细节的基础上, 考虑到更广泛的上下文信息, 在处理长条状阴影时分割效果更好。

表 2 Vaihingen 数据集上的对比实验结果

Method	Per-class F1					MeanF1	OA	mIoU
	Imp.Sur.	Building	Low Veg.	Tree	Car			
UNet ^[9]	89.86	93.41	80.82	86.93	81.50	86.51	87.79	76.55
PSPNet ^[30]	91.38	94.20	83.05	88.71	75.02	86.47	89.36	76.78
UNetFormer ^[15]	92.70	95.30	84.90	90.60	88.50	90.40	91.20	82.70
FDNet ^[18]	92.87	95.85	85.04	90.35	88.14	90.45	91.59	82.72
Xnet ^[31]	92.89	95.48	85.17	91.10	88.22	90.57	91.21	82.95
UperNet-FDAC ^[32]	92.80	95.78	85.50	90.83	89.40	90.86	91.10	83.49
MIFNet ^[22]	93.59	96.13	86.34	91.28	90.06	91.48	91.90	84.53
AFANet(本文)	94.15	96.02	86.92	91.80	90.24	91.83	93.15	85.13

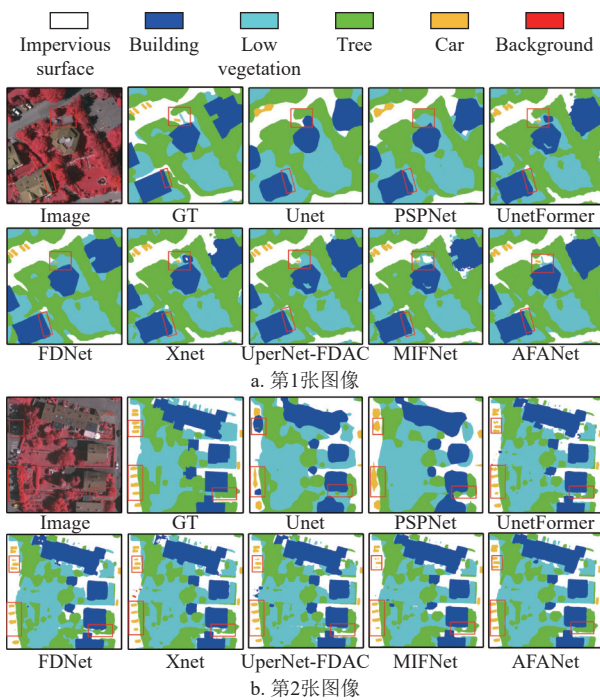


图 5 Vaihingen 数据集上的可视化对比图

2.3 消融实验

为了验证 AFANet 网络中各模块的有效性, 在两个数据集上进行了消融实验, 具体数据如表 3 所示, 可视化结果见图 6。

表 3 AFANet 消融实验结果

Dataset	Method	%		
		MeanF1	OA	mIoU
Vaihingen	Baseline(ConvNext)	90.21	92.38	82.49
	Baseline+FDPM	91.01	92.71	83.26
	Baseline+FDPM+DDLb	91.63	92.94	84.76
	Baseline+FDPM+DDLb+DEM	91.83	93.15	85.13
	Baseline(ConvNext)	92.15	91.08	85.69
Potsdam	Baseline+FDPM	92.92	91.53	86.78
	Baseline+FDPM+DDLb	93.15	92.10	87.56
	Baseline+FDPM+DDLb+DEM	93.38	92.35	87.81

1) Baseline+FDPM。基准模型 (Baseline) 选用 ConvNext-tiny 作为主干网络, 在跳跃连接处加

入频率动态融合模块 (FDPM) 后性能均有提升。在 Vaihingen 数据集上 F1 分数提升了 0.8%, 总体精度 (OA) 提升了 0.33%, 平均交并比 (mIoU) 提升了 0.77%; 在 Potsdam 数据集上 F1 分数提升了 0.77%, 总体精度 (OA) 提升了 0.45%, 平均交并比 (mIoU) 提升了 1.09%。第 1 幅图中与楼房颜色相近的道路和第 2 幅图中的树木通过高低频信息的融合实现了有效分割。

2) Baseline+FDPM+DDLb。加入双域学习模块后, AFANet 的性能进一步提升。在 Vaihingen 数据集上 F1 分数、总体精度 (OA) 和平均交并比 (mIoU) 分别达到了 91.63%、92.94% 和 84.76%。在 Potsdam 数据集上表现更为出色, 各项指标达到了 93.15%、92.10% 和 87.56%。第 1 幅图中的低矮植被与第 3 幅图中的房屋, 都属于难以识别的细长对象。AFANet 通过双域学习模块中的高低频双分支学习, 最大程度避免了低频平滑操作对细节的破坏, 保留轮廓细节。并且采用了多尺度条形卷积, 增强了横向与竖向特征的连续性, 有效解决细长对象的分割难题。另外双域学习模块通过局部分支的通道空间注意力学习, 弥补了局部细节的丢失, 完善了房屋建筑等对象的轮廓细节。

3) Baseline+FDPM+DDLb+DEM。细节增强模块的引入, 可以有效解决光照条件导致的阴影问题。4 幅图中存在不少被光照阴影遮挡的汽车, 加入细节增强模块后, 汽车的轮廓被进一步精确识别, 使得 AFANet 受到阴影的影响更小。在第 3 幅图中, 有 4 个被阴影遮挡的整齐排列的箱子, 通过前两个模块利用频率的学习易被分为汽车。细节增强模块利用差分卷积的梯度增强, 有效去除了阴影对箱子特征的干扰, 恢复了箱子棱角等被掩盖的细节。细节增强模块使 AFANet 在复杂光照场景中展现出更强的鲁棒性, 避免了“以形误判”的局

限性。

通过特征图的可视化可以验证自适应低通和高通滤波器提取的高低频信息对于结果的影响。如图 7 所示, 可以观察到, 低频特征图含有丰富的区域语义信息, 不同深浅的颜色代表模型对不同区域

语义划分的理解。高频特征图则包含大量的细节描述, 如建筑、汽车或其他对象的一些轮廓、纹理特征。频率动态融合模块将两部分信息分别生成并融合, 融合后的特征图实现了高低频信息的互补, 有效提高了分割的精度。

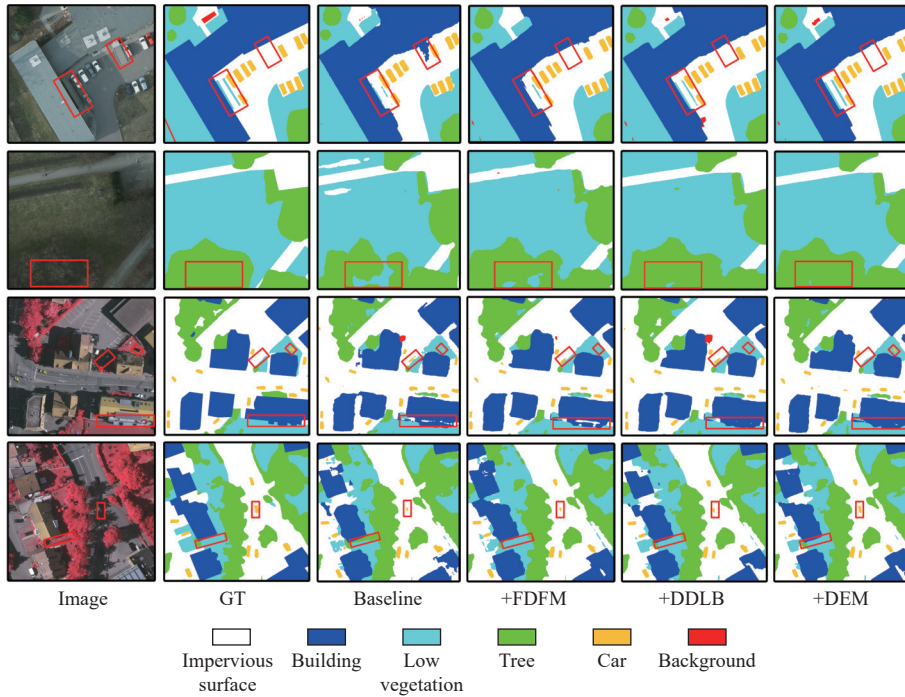


图 6 消融可视化对比图

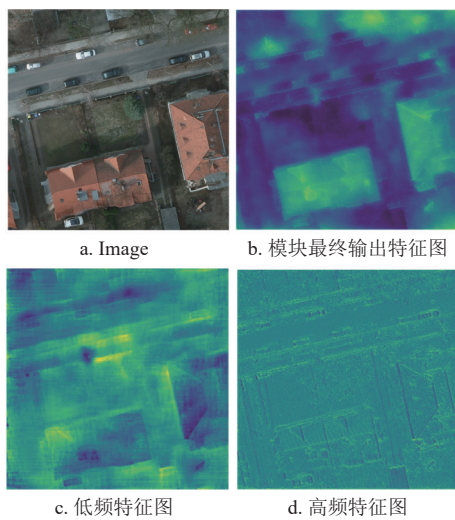


图 7 频率动态融合模块特征图可视化

图 8 选取了 3 种具有代表性的类别: 汽车、低矮植被和树木, 在 Vaihingen 测试集上绘制了 Grad-CAM 热力图用来展示各模块对关键区域的特征聚焦能力。可以看到对于低矮植被和树木两个类别基准模型只能将大体轮廓分清, 通过引入频率动态融

合模块和双域学习模块后, 模型可以利用频率信息更好地对这两种极为相似的类别加以区分。激活区域进一步扩大, 对目标区域的关注度呈现显著提升, 表现出更强的梯度响应。而加入细节增强模块后, 在光照阴影下的汽车开始受到关注, 有效减轻了阴影遮挡的干扰, 提高了模型的性能。

为了验证频率动态融合模块中所提自适应低通和高通滤波频率分析方法的优越性, 与几种常见的频率分析方法进行对比, 结果如表 4 所示。传统的傅里叶变换是全局操作无法捕捉局部时频特性, 并且高低频信息混合处理对于非平稳特征无法动态捕获。小波变换虽能将低频分量进行解耦但受限于固定基函数, 对噪声敏感的特性可能会使高频噪声被放大。全局平均池化是一种简便提取低频分量的方法, 但本质上是一种固定低通滤波, 而遥感图像中不同区域的低频结构差异显著, 固定策略难以适配这种多样性, 过度平滑导致局部结构丢失, 影响分割精度。本文提出的自适应低通和高通滤波能有效规避上述问题, 并且取得了良好的实验结果。

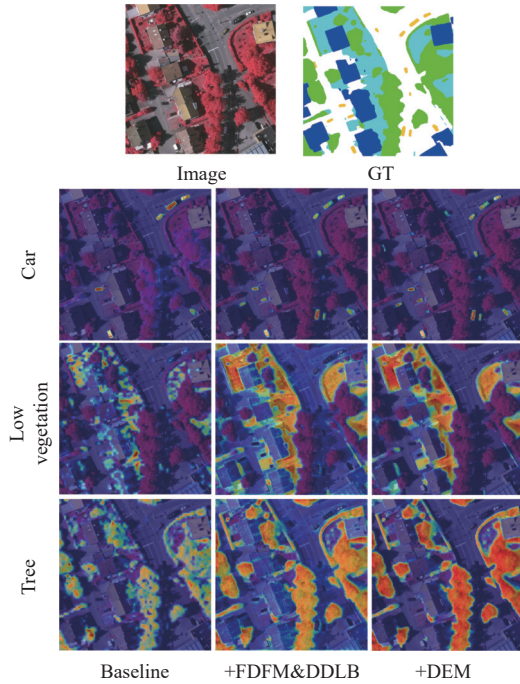


图 8 各模块 Grad-CAM 热力图

表 4 不同频率分析方法的对比结果

Dataset	Method	MeanF1	OA	mIoU
Potsdam	Fourier Transform	91.61	89.82	85.72
	Wavelet Transform	92.58	90.76	86.51
	Average Pool	92.36	90.64	86.02
	FDFM(本文)	93.38	92.35	87.81

为了验证本文使用的损失函数对 AFANet 性能的影响,进行损失函数对比实验,结果如表 5 所示。交叉熵损失容易忽略类别不平衡问题。引入 Dice 损失后,通过其重叠优化机制,提升分割结果的空间连续性与类不平衡鲁棒性,同时增强对小目标的梯度敏感性。边界损失采用拉普拉斯卷积与阈值划分的方法来监督模型在边界区域的表现,更侧重于局部边缘特征,而容易忽略整体结构的信息,对整体性能的提升作用有限。辅助损失通过约束核心频率动态融合模块,从而增强其对不同图像内容的频率特征自适应调整能力,提升模型在复杂场景下的语义一致性与细节补偿。

针对提出的通道空间协同模块,表 6 的对比实验结果表明其在遥感图像分割任务中展现出显著优势。在经典注意力模块的性能对比分析中, CBAM 模块因依赖全局平均池化操作引发高频特征丢失,导致小尺度目标(如建筑轮廓及植被纹理)的细节表征能力降低。ECA 模块受限于单一通道注意力架构,缺乏空间维度特征校准,难以有效捕捉遥感图像中具有方向特性的线状目标(如道路网络及条

带状植被)。Shuffle Attention 模块虽采用并行双分支结构融合通道与空间特征,但其独立计算方式导致跨维度注意力交互不足,可能引发通道特征强化与空间区域关注间的协同性不足问题。通道空间协同模块可以有效弥补以上问题且在多项客观评价指标上均取得最优性能表现。

表 5 不同损失函数的对比结果

Dataset	Main Loss			Aux Loss	mIoU
	Cross Entropy	Dice	Edge	Cross Entropy	
Potsdam	√	—	—	—	86.45
	√	√	—	—	87.43
	√	—	√	—	86.75
	√	√	—	√	87.81

表 6 不同注意力模块的对比结果

Dataset	模块	MeanF1	OA	mIoU
Potsdam	+ CBAM ^[33]	93.26	92.24	87.71
	+ ECA ^[34]	93.14	91.95	87.63
	+ Shuffle Attention ^[35]	93.23	92.17	87.62
	+ CSSM(本文)	93.38	92.35	87.81

3 结束语

针对现有遥感图像分割方法在特征融合时缺乏对低频结构与高频细节的协同处理,同时难以对图像内容自适应响应的问题,提出并实现了一种自适应频率感知网络。在跳跃连接处构建频率动态融合模块进行跨尺度特征融合,通过自适应高通与低通滤波器的协同作用,收集高分辨率编码器的高频特征和低分辨率解码器的低频特征。通过双域学习模块,集成局部空间和全局频域信息。在实验环节中,通过定量分析与可视化手段验证了网络的有效性,并在公共数据集上取得了最优表现,相较于传统方法,自适应频率感知网络在边界精度与类内一致性上表现更优,分割结果更具可靠性。

未来的工作将致力于在维持模型性能的前提下,降低其复杂度,以更少的参数高效学习图像特征。同时,为了增强模型的泛化能力,将进一步探索其在更多不同数据集上的表现,旨在不断推进遥感图像语义分割领域的发展。

参考文献

- [1] VICTOR N, MADDIKUNTA P K R, MARY D R K, et al. Remote sensing for agriculture in the era of industry 5.0—A survey[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 5920-5945.

- [2] KEMARAU R A, SUAB S A, EBOY O V, et al. Integrative approaches in remote sensing and GIS for assessing climate change impacts across Malaysian ecosystems and societies[J]. *Sustainability*, 2025, 17(4): 1344.
- [3] JIA P, CHEN C, ZHANG D, et al. Semantic segmentation of deep learning remote sensing images based on band combination principle: Application in urban planning and land use[J]. *Computer Communications*, 2024, 217: 97-106.
- [4] WENQI Y U, GONG C, MEIJUN W, et al. MAR20: A benchmark for military aircraft recognition in remote sensing images[J]. *National Remote Sensing Bulletin*, 2024, 27(12): 2688-2696.
- [5] WU Z, GAO Y, LI L, et al. Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold[J]. *Connection Science*, 2019, 31(2): 169-184.
- [6] MARMANIS D, SCHINDLER K, WEGNER J D, et al. Classification with an edge: Improving semantic image segmentation with boundary detection[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 135: 158-172.
- [7] 张晓磊, 潘卫军, 王思禹, 等. 一种基于双边模糊聚类的遥感图像分割算法[J]. *航空计算技术*, 2020, 50(1): 33-37.
- ZHANG X L, PAN W J, WANG S Y, et al. A remote sensing image segmentation algorithm based on bilateral fuzzy clustering[J]. *Aeronautical Computing Technique*, 2020, 50(1): 33-37.
- [8] 马妍, 古丽米拉克孜别克. 图像语义分割方法在高分辨率遥感影像解译中的研究综述[J]. *计算机科学与探索*, 2023, 17(7): 1526-1548.
- MA Y, KEZIERBIEKE G. A review of image semantic segmentation methods in high-resolution remote sensing image interpretation[J]. *Journal of Computer Science and Exploration*, 2023, 17(7): 1526-1548.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//*Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer International Publishing, 2015: 234-241.
- [10] DIAKOIANNIS F I, WALDNER F, CACCETTA P, et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 162: 94-114.
- [11] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. [2025-03-12]. <https://arxiv.org/pdf/2010.11929>.
- [12] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 10012-10022.
- [13] ZHENG S X, LU J C, ZHAO H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2021: 6881.
- [14] ZHANG R H, ZHANG Q, ZHANG G X. LSRFormer: Efficient transformer supply convolutional neural networks with global information for aerial image segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-13.
- [15] 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.
- [16] WU H L, HUANG P, ZHANG M, et al. CMTFNet: CNN and multiscale transformer fusion network for remote-sensing image semantic segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-12.
- [17] ZHANG J, SHAO M W, WAN Y C, et al. Boundary-aware spatial and frequency dual-domain transformer for remote sensing urban images segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5637718.
- [18] LIU H J, ZHOU X Y, WANG C L, et al. Fourier-deformable convolution network for road segmentation from remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 4415117.
- [19] DUAN S, ZHAO J, HUANG X, et al. Semantic segmentation of remote sensing data based on channel attention and feature information entropy[J]. *Sensors*, 2024, 24(4): 1324.
- [20] YANG Y S, YUAN G J, LI J J. SFFNet: A wavelet-based spatial and frequency domain fusion network for remote sensing segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 3000617.
- [21] DONG B, WANG P C, WANG F. Head-free lightweight semantic segmentation with linear transformer[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 37(1): 516-524.
- [22] FAN J Y, LI J J, LIU Y P, et al. Frequency-aware robust multidimensional information fusion framework for remote sensing image segmentation[J]. *Engineering Applications of Artificial Intelligence*, 2024, 129: 107638.
- [23] WOO S, DEBNATH S, HU R H, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 16133-16142.
- [24] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-

- aware reassembly of features[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 3007-3016.
- [25] ABDEL M S, ZHANG Y L, WEI D L, et al. Dynamic high-pass filtering and multi-spectral attention for image super-resolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 4288-4297.
- [26] CHEN L, FU Y, GU L, et al. Frequency-aware feature fusion for dense image prediction[J]. *IEEE Trans Pattern Anal Mach Intell*, 2024, 46(12): 10763-10780.
- [27] SU Z, LIU W Z, YU Z T, et al. Pixel difference networks for efficient edge detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 5117-5127.
- [28] XU J F, BODDETI V N, SAVVIDES M. Local binary convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 19-28.
- [29] CHEN Z, HE Z, LU Z M. DEA-net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. *IEEE Trans Image Process*, 2024, 33: 1002-1015.
- [30] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 2881.
- [31] ZHOU Y F, HUANG J X, WANG C L, et al. XNet: Wavelet-based low and high frequency fusion networks for fully- and semi-supervised semantic segmentation of biomedical images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Vancouver: IEEE, 2023: 21085-21096.
- [32] CHEN L, GU L, ZHENG D, et al. Frequency-adaptive dilated convolution for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 3414-3425.
- [33] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Munich: Springer, 2018: 3-19.
- [34] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11534-11542.
- [35] ZHANG Q L, YANG Y B. Sa-net: Shuffle attention for deep convolutional neural networks[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing. Melbourne: IEEE, 2021: 2235-2239.

责任编辑 税 红