引文格式:孙根云,苗静茹,赵元昊,等.融合细节-全局特征与结构增强的多源遥感影像点注释弱监督语义分割[J].武汉大学学报(信息科学版),2026,51(11):2424-2435.DOI:10.13203/j.whugis20260280

Citation: SUN Genyun, MIAO Jingru, ZHAO Yuanhao, et al. Point-Annotation-Based Weakly Supervised Semantic Segmentation of Multi-source Remote Sensing Images via Detail-Global Feature Fusion and Structure Enhancement[J]. Geomatics and Information Science of Wuhan University, 2026, 51(11): 2424-2435. DOI: 10.13203/j.whugis20260280

融合细节-全局特征与结构增强的多源遥感影像点注释弱监督语义分割

孙根云[1,2] 苗静茹[2] 赵元昊[2] 张爱竹[2] 张文娟[3]
马 毅[4] 孙 林[5]

1 自然资源部华南热带亚热带自然资源监测重点实验室,广东 广州,510663
2 中国石油大学(华东)海洋与空间信息学院,山东 青岛,266580
3 中国科学院空天信息创新研究院,北京,100094
4 自然资源部第一海洋研究所,山东 青岛,266061
5 山东科技大学测绘与空间信息学院,山东 青岛,266590

摘 要:点注释弱监督能够显著降低遥感影像语义分割的像素级标注成本,但由于监督信号高度稀疏,模型容易出现类别语义表达不足、目标边界模糊及细小地物遗漏等问题。与此同时,复杂城市区域地物类型多样、尺度差异明显,单源遥感影像难以充分表征地物的光谱、纹理和空间结构特征。因此,融合多源遥感影像的互补信息已成为提升复杂场景语义分割性能的重要方向。然而,现有多源弱监督方法对局部细节互补、全局语义交互和边界结构约束的联合建模仍不充分。针对上述问题,提出一种融合细节-全局特征与结构增强网络(detail-global feature fusion and structure enhancement network,DGSE-Net),用于多源弱监督遥感影像语义分割。DGSE-Net由多尺度特征提取、尺度自适应特征组合以及解码与输出3个部分组成。其中,多尺度特征提取部分采用分割大模型(segment anything model,SAM)的图像编码器和ResNet18编码器,分别提取主源数据与辅助源数据的多尺度特征;尺度自适应特征组合部分首先通过细节-全局多源特征融合(detail-global multi-source feature fusion,DGCF)模块在不同尺度上独立建模局部空间细节互补与全局语义交互,提高多源特征的一致性与判别能力,然后通过结构感知边缘增强(structure-aware edge enhancement,SAEE)模块引入Sobel边缘先验、邻域权重降噪、浅层局部卷积和深层窗口注意力,强化弱监督条件下地物边界、空间结构与细小目标的特征表达;解码与输出部分采用跨分辨率特征集成(cross-resolution feature integration,CRFI)解码模块,对多尺度增强特征进行逐步聚合,恢复空间细节并生成语义分割结果。在ISPRS Potsdam和ISPRS Vaihingen数据集上的点注释弱监督实验表明,DGSE-Net在仅使用0.01%像素点标签的条件下取得了优于对比方法的分割精度,消融实验进一步验证了DGCF、SAEE和CRFI的有效性。DGSE-Net能够在极稀疏点监督条件下有效利用多源遥感数据的互补信息,为低标注成本的高分辨率遥感影像精细语义分割提供了一种有效方法。

关键词:弱监督语义分割;多源遥感影像;点注释;特征融合;结构增强

中图分类号:P237 **文献标识码**:A **收稿日期**:2026-07-24

DOI:10.13203/j.whugis20260280 **文章编号**:1671-8860(2026)11-2424-12

Point-Annotation-Based Weakly Supervised Semantic Segmentation of Multi-source Remote Sensing Images via Detail-Global Feature Fusion and Structure Enhancement

SUN Genyun[1,2] *MIAO Jingru*[2] *ZHAO Yuanhao*[2] *ZHANG Aizhu*[2] *ZHANG Wenjuan*[3]
MA Yi[4] *SUN Lin*[5]

1 Key Laboratory of Natural Resources Monitoring in Tropical and Subtropical Area of South China, Ministry of Natural Resources,

基金项目:自然资源部华南热带亚热带自然资源监测重点实验室开放基金(2024NRMK03);国家自然科学基金(42271347,42371350);山东省自然科学基金(ZR2024MD095, ZR2024MD030)。

第一作者:孙根云,博士,教授,研究方向为多源遥感数据智能处理。genyunsun@163.com

通信作者:张爱竹,博士,副教授。zhangaizhu789@163.com

Guangzhou 510663, China
2 College of Ocean and Space Information, China University of Petroleum (East China), Qingdao 266580, China
3 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
4 The First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China
5 College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

Abstract: **Objectives**: Point-annotation-based weak supervision can substantially reduce the cost of pixel-level labeling for remote sensing image semantic segmentation. However, the supervision signals provided by point labels are extremely sparse, which often leads to insufficient category-level semantic representation, blurred object boundaries, and omission of small ground objects. Meanwhile, complex urban scenes contain diverse land-cover categories with significant scale variations, irregular spatial distributions, and intricate boundary structures. Single-source remote sensing imagery is usually insufficient to fully characterize the spectral, textural, and spatial structural features of different ground objects. Therefore, integrating complementary information from multi-source remote sensing images has become an important direction for improving semantic segmentation performance in complex scenarios. Nevertheless, the existing multi-source weakly supervised methods still lack sufficient joint modeling of local detail complementarity, global semantic interaction, and boundary structure constraints. To address these challenges, this paper proposes a detail-global feature fusion and structure enhancement network (DGSE-Net) for multi-source weakly supervised remote sensing image semantic segmentation. **Methods**: DGSE-Net consists of three main parts, including multi-scale feature extraction, scale-adaptive feature combination, and decoding output. First, in the multi-scale feature extraction, the image encoder of segment anything model (SAM) and a ResNet18 encoder are employed to extract multi-scale features from the primary source data and auxiliary source data, respectively. Second, in the scale-adaptive feature combination, a detail-global multi-source feature fusion (DGCF) module first models local spatial detail complementarity and global semantic interaction independently at different scales, thereby improving the consistency and discriminative capability of multi-source features. Then, a structure-aware edge enhancement (SAEE) module introduces Sobel edge priors, neighborhood-weighted denoising, shallow local convolution, and deep window attention to strengthen the representation of object boundaries, spatial structures and small targets under weak supervision. Third, in the decoding output, a cross-resolution feature integration (CRFI) decoding module is adopted to progressively aggregate the enhanced multi-scale features, recover spatial details, and generate semantic segmentation results. Point-annotation-based weakly supervised experiments are conducted on the ISPRS Potsdam and ISPRS Vaihingen datasets. **Results**: Using only 0.01% pixel-level point labels, DGSE-Net achieves higher segmentation accuracy than the comparison methods. Ablation experiments further verify the effectiveness of DGCF, SAEE, and CRFI, demonstrating that local detail enhancement, global cross-source semantic interaction, structure-aware edge enhancement, and cross-resolution feature integration all contribute to performance improvement. **Conclusions**: The experimental results indicate that DGSE-Net can effectively exploit complementary information from multi-source remote sensing data under extremely sparse point supervision, providing an effective solution for fine-grained semantic segmentation of high-resolution remote sensing images with low annotation cost.

Key words: weakly supervised semantic segmentation; multi-source remote sensing images; point annotation; feature fusion; structure enhancement

遥感影像语义分割旨在为影像中每个像素赋予对应的地物类别标签,是城市精细制图、土地覆盖调查、生态环境监测和灾害评估等任务的重要基础[1-3]。随着高分辨率、多传感器遥感数据获取能力的不断增强,遥感影像能够更加精细地刻画城市地表空间结构和地物分布特征。然而,复杂城市区域中的建筑、道路、植被、车辆等目标通常呈现出尺度差异大、边界形态复杂、类间纹理相似和类内变化显著等特点,使得遥感影像精细化智能解译仍面临较大挑战[4-7]。

近年来,深度学习方法凭借强大的特征表达能力和端到端学习机制,显著推动了遥感影像语义分割的发展。早期全卷积网络、编码器-解码器结构和空洞卷积等方法为遥感影像语义分割奠定了基础[8-10]。在此基础上,研究者们从单纯依赖卷积局部感受野转向结合全局上下文建模、多尺度特征交互和大模型先验知识的精细化分割框架。例如,文献[11]将Transformer引入遥感影像语义分割,通过自注意力机制增强长距离依赖建模能力;文献[12]进一步面向多模态遥感场景构建多层级特征交互结构,提升多源遥感数据在多尺度层面的协同表达能力;文献[13]围绕遥感基础模型开展系统研究,指出大规模预训练模型在稀疏标注任务中的潜力。与此同时,分割大模型(segment anything model,SAM)及其遥感适配方法也为低标注成本分割提供了新的思路。文献[14]利用SAM生成更完整的伪标签或超像素区域,以缓解弱监督分割中目标检测不完整和边界定位不准确的问题;文献[15]进一步探索点注释条件下的SAM微调策略,为点监督遥感影像分割提供了新的技术路径。上述研究表明,遥感影像语义分割正由传统全监督单源建模,向多源协同、弱监督学习和基础模型辅助的方向发展。

尽管全监督深度分割模型在遥感场景中已取得较大进展,但其性能通常依赖大量高质量像素级标注样本。高分辨率遥感影像覆盖范围广、地物类别复杂、目标边界细碎,逐像素标注不仅耗时耗力,而且对标注人员的专业知识要求较高,严重限制了深度分割模型在大范围遥感制图任务中的推广应用。为降低标注成本,弱监督语义分割方法逐渐受到关注。与图像级标签、框标注和涂鸦标注等弱监督形式相比,点注释仅需为少量代表性像素赋予类别标签,具有标注效率高、实施成本低和空间定位明确等优点。因此,近年来,已有研究尝试利用点标签、伪标签优化、一致性约束和大模型先验提升遥感弱监督分割性能[16-19]。然而,点标签只提供离散位置上的类别监督,难以直接约束完整地物区域和精细边界,模型训练过程中容易出现伪监督不足、类别扩散、边界混淆以及细小目标漏分等问题[15,20-22]。

多源遥感数据为缓解弱监督条件下语义信息不足提供了新的思路。主源影像通常包含较为丰富的光谱颜色、视觉纹理和空间轮廓信息,能够有效表征地物的表观特征;辅助源数据则可从不同观测角度补充地物结构、几何形态或物理响应差异,有助于提升复杂场景下的类别可分性[23-28]。近年来,多源遥感语义分割研究更加关注跨源特征对齐、层级语义交互和模态差异抑制[29]。文献[30]综述了光学与合成孔径雷达(synthetic aperture radar,SAR)图像深度特征融合在语义分割中的研究进展;文献[12]构建多层级多模态融合Transformer,通过卷积神经网络(convolutional neural network,CNN)与视觉Transformer协同建模,提升不同源数据在多尺度层面的融合表达能力;文献[31]进一步将SAM大模型先验引入多模态遥感语义分割,通过多模态微调框架和适配器结构增强不同模态特征的有效利用。然而,在点注释弱监督场景下,多源特征利用仍存在明显困难。一方面,极稀疏监督信号难以为不同数据源之间的特征对齐和语义互补提供充分约束;另一方面,不同源数据在空间分辨率、纹理分布、成像机理和语义响应上存在差异,直接拼接或简单相加难以判断不同区域应依赖哪一来源特征,还可能引入冗余响应、局部噪声和跨源干扰。

针对上述研究问题与难点,本文提出一种融合细节-全局特征与结构增强网络(detail-global feature fusion and structure enhancement network,DGSE-Net)。DGSE-Net以点注释为监督形式,通过同时关注局部细节差异与全局语义关联,并进一步引入结构边缘信息强化弱监督条件下的地物语义信息,最终实现弱监督条件下的多源遥感影像语义分割。

1 原理与方法

1.1 网络总体结构

DGSE-Net整体由多尺度特征提取、尺度自适应特征组合、解码与输出3个部分组成,整体结构如图1所示。在多尺度特征提取部分,网络以光学影像作为主源输入,以数据集配套的数字表面模型(digital surface model,DSM)作为辅助源输入。其中,SAM图像编码器用于提取光学影像的多尺度特征,ResNet18编码器用于提取对应DSM的多尺度结构与高程响应特征,为后续多源特征融合提供基础。在尺度自适应特征组合部分,网络首先通过3个细节-全局多源特征融合(detail-global multi-source feature fusion,DGCF)模块分别建模局部细节互补与全局语义交互。3个尺度的DGCF模块输出采用自底向上的逐级聚合方式,首先将32×32×256的深层融

合特征上采样 2 倍,并与 64×64×256 的中间尺度融合特征逐元素相加,得到中间尺度聚合特征;然后将该中间尺度聚合特征再次上采样 2 倍,并与 128×128×256 的浅层融合特征逐元素相加,得到浅层聚合特征。随后,融合特征通过特征聚合逐步整合深层语义信息与浅层空间细节,并输入到结构边缘增强(structure-aware edge enhancement, SAEE)模块,进一步强化弱监督条件下地物边界、细小目标和复杂结构区域的表达能力。在解码与输出部分,不同尺度的结构增强特征共同输入跨分辨率特征集成(cross-resolution feature integration,CRFI)解码模块[20]。该模块在以往研究中已经被证明可以对多尺度特征进行跨分辨率集成与解码,充分融合深层语义信息和浅层空间细节,最终输出与输入影像尺寸一致的语义分割结果。

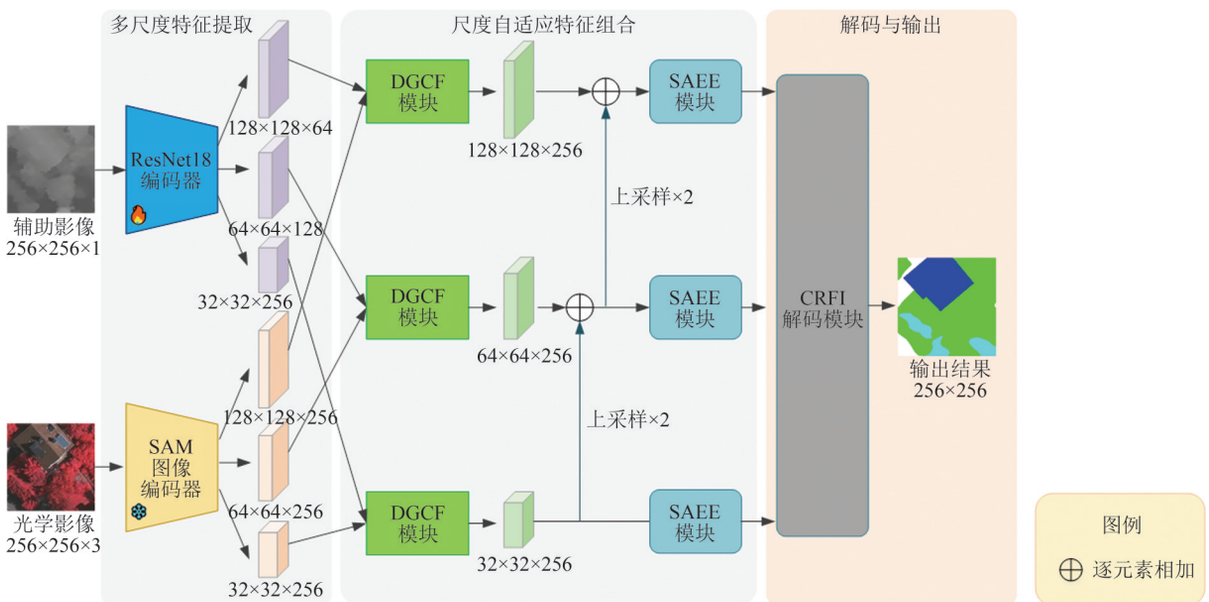


图 1　DGSE-Net 网络总体结构

Fig. 1　Overall Architecture of DGSE-Net

1.2　多源特征编码器

多源遥感数据通常具有不同的空间响应和语义表达方式。如果直接将不同来源数据在输入端拼接,网络容易将不同来源数据中的有效信息、冗余响应和局部干扰混合在一起,从而影响后续特征学习。为此,本文采用多分支编码结构分别提取不同来源数据的特征。

具体而言,将光学影像与同一区域、同一空间分辨率的 DSM 分别送入两个独立编码器。主源分支采用 SAM 图像编码器提取光谱、纹理与空间轮廓特征;辅助源分支采用 ResNet18 编码器提取 DSM 中的高程起伏、地物几何形态与空间结构特征。编码器采用差异化设计是因为 SAM 图像编码器具有较强的视觉结构感知能力,但是 SAM 的预训练数据大多是光学等被动成像的数据,因此适合从主源光学遥感数据中提取边界、形状和上下文特征;但是如果将 SAM 图像编码器应用在 DSM、SAR 等主动成像遥感图像中,反而会放大噪声,污染分割结果。而 ResNet18 结构相对轻量,能够以较低复杂度从辅助源数据中提取局部纹理和层级结构特征,因此,更适合用于辅助源数据的多尺度特征提取。

与单一分支编码相比,多分支编码具有两个优势。首先,不同来源数据在进入融合阶段之前可以分别完成特征抽象,从而减少源间差异对早期特征学习的干扰。其次,多尺度特征为后续融合提供了更加丰富的信息层次,使模型既能够利用浅层空间细节,也能够利用深层语义信息。因此,本文在多个尺度上对不同来源特征进行融合,而不是仅在编码器末端进行一次融合,这有助于提高模型对复杂地物结构和尺度变化的适应能力。

1.3　细节-全局多源特征融合模块

在多源弱监督遥感影像语义分割中,点注释仅能为少量像素位置提供类别监督,难以充分约束完整地物区域、精细边界和跨源语义关系。因此,特征融合模块需要在有限监督条件下主动挖掘主源与辅助源之间的互补信息,同时抑制辅助源中可能存在的局部噪声和冗余响应。为此,本文设计 DGCF 模块,其结构如图 2 所示。

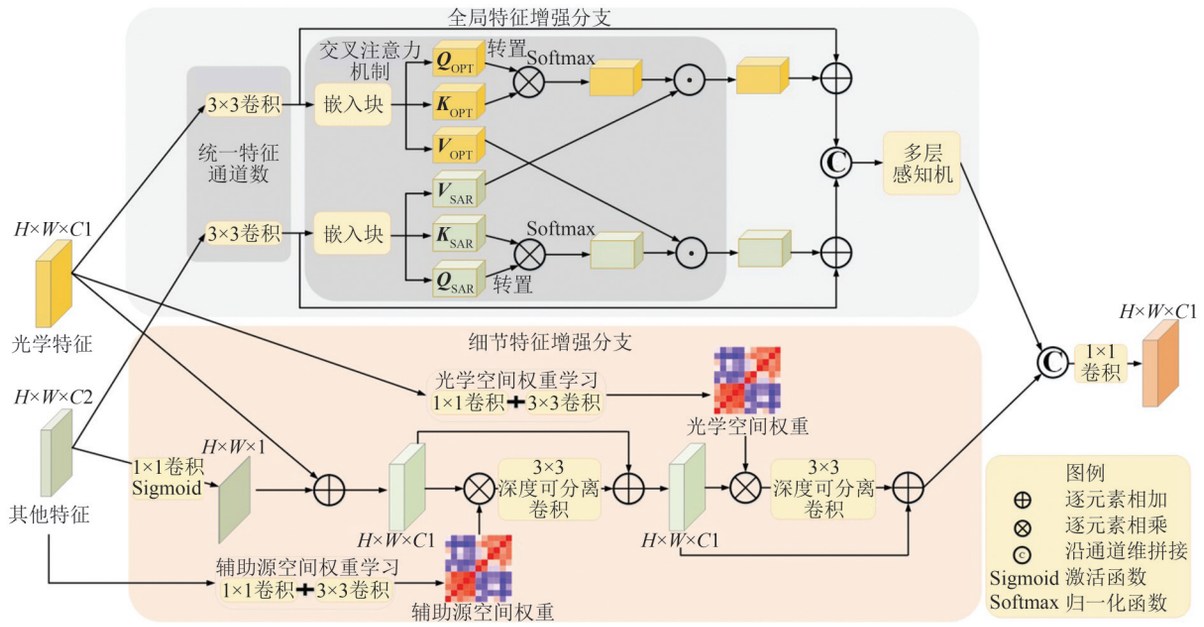


图2 细节-全局多源特征融合模块

Fig. 2 Detail-Global Multi-source Feature Fusion Module

DGCF独立接收同尺度的主源特征与辅助源特征，通过细节特征增强分支和全局特征增强分支分别建模局部结构互补与长距离语义交互，最终输出多尺度的多源融合特征。

细节特征增强分支主要用于强化局部边界、纹理细节和细小目标表达。由于主源影像通常具有较清晰的视觉纹理和空间轮廓，而辅助源数据能够从另一观测角度补充地物结构差异，二者在局部区域具有较强互补性。具体来说，该分支首先利用1×1卷积和Sigmoid激活从辅助源特征中生成辅助源引导权重，并将其用于调制辅助源响应，使辅助源信息能够以权重形式参与主源特征增强；然后将主源特征与加权后的辅助源特征进行融合，得到初步组合特征。在此基础上，分别构建主源空间权重生成单元和辅助源空间权重生成单元，通过1×1卷积与3×3卷积学习不同空间位置的可靠性响应。其中，主源空间权重用于突出纹理清晰、边界明显和类别响应稳定的区域；辅助源空间权重用于补充局部结构差异。在获得空间权重后，细节特征增强分支通过逐级权重调制和局部卷积增强组合特征。首先，利用降噪后的辅助源空间权重对组合特征进行加权，引导模型关注辅助源中结构响应较可靠的区域，并通过3×3深度可分离卷积提取轻量化局部结构信息；然后将增强结果与前级特征进行残差融合，以保留原始主源表达并避免弱监督条件下的细节退化。进一步地，主源空间权重被用于再次调制特征响应，并结合3×3深度可分离卷积强化主源中的边界纹理和局部几何信息。

尽管细节特征增强分支能够有效补充局部纹理和边界信息，但遥感影像中的建筑群、道路、水体和植被等地物往往具有明显的空间连续性和上下文依赖关系，仅依赖局部特征容易造成区域语义不一致或类间混淆。为此，DGCF进一步构建全局特征增强分支，用于建模主源与辅助源之间的长距离语义关联。如图2所示，该分支首先对输入特征进行大尺度池化和3×3卷积处理，在降低计算复杂度的同时保留主要空间结构信息；然后将主源特征与辅助源特征划分为嵌入块，并分别映射为查询向量$\boldsymbol{Q}$、键向量$\boldsymbol{K}$和值向量$\boldsymbol{V}$，下标OPT和SAR分别表示光学影像特征和SAR影像特征。通过双向跨源注意力交互，主源特征能够引入辅助源中的结构上下文，辅助源特征也能够获得主源中的纹理与语义信息，从而在全局范围内建立跨源语义对应关系。具体来说，全局特征增强分支分别计算主源和辅助源的注意力响应，并通过Softmax获得全局关系权重。随后，利用跨源值特征进行信息交互，使两类特征在长距离范围内实现语义补充。交互后的特征分别与对应输入特征进行残差融合，以保持原始语义表达并增强全局一致性。最后，双向交互结果进行拼接，并经过多层感知机或通道映射完成特征整合，得到全局增强特征。这能够帮助模型缓解点注释监督稀疏导致的区域语义不连续问题，使模型在缺少密集标签约束的情况下仍能利用全局上下文推断地物类别关系。

最后,DGCF 模块将细节特征增强分支与全局特征增强分支的输出进行拼接,并通过1×1卷积完成通道压缩与特征整合,输出多尺度的多源融合特征。与直接拼接或简单相加的多源融合方式相比,DGCF 使局部边界纹理信息与全局语义上下文在同一尺度内实现互补融合。其中,细节特征增强分支提供更加稳定的局部结构响应,全局特征增强分支补充跨源长距离语义关联。

1.4　结构边缘增强模块

在点注释弱监督语义分割中,监督信号仅分布于少量像素位置,难以直接约束完整地物轮廓和类别交界区域,容易导致建筑边缘、道路边界以及车辆等细小目标出现粘连、断裂或漏分等问题。为增强弱监督条件下的结构感知能力,本文构建SAEE模块,其结构如图3所示。

具体来说,SAEE模块通过噪声响应抑制、显式边缘引导、尺度自适应上下文建模3种方式展开特征增强。首先,构建降噪分支,削弱局部不稳定响应对结构增强的干扰。该分支通过卷积映射提取输入特征的局部邻域响应,并根据中心像素与邻域像素之间的响应差异计算邻域权重。权重较高的位置通常表示与中心像素具有更强的一致性,而差异较大的位置则可能对应噪声、伪边界或不稳定响应。然后,利用归一化后的邻域权重对局部特征进行加权聚合,得到噪声抑制后的结构响应特征。该过程能够在弱监督标签稀疏的情况下增强区域内部一致性,并降低局部噪声对后续边缘增强的影响。

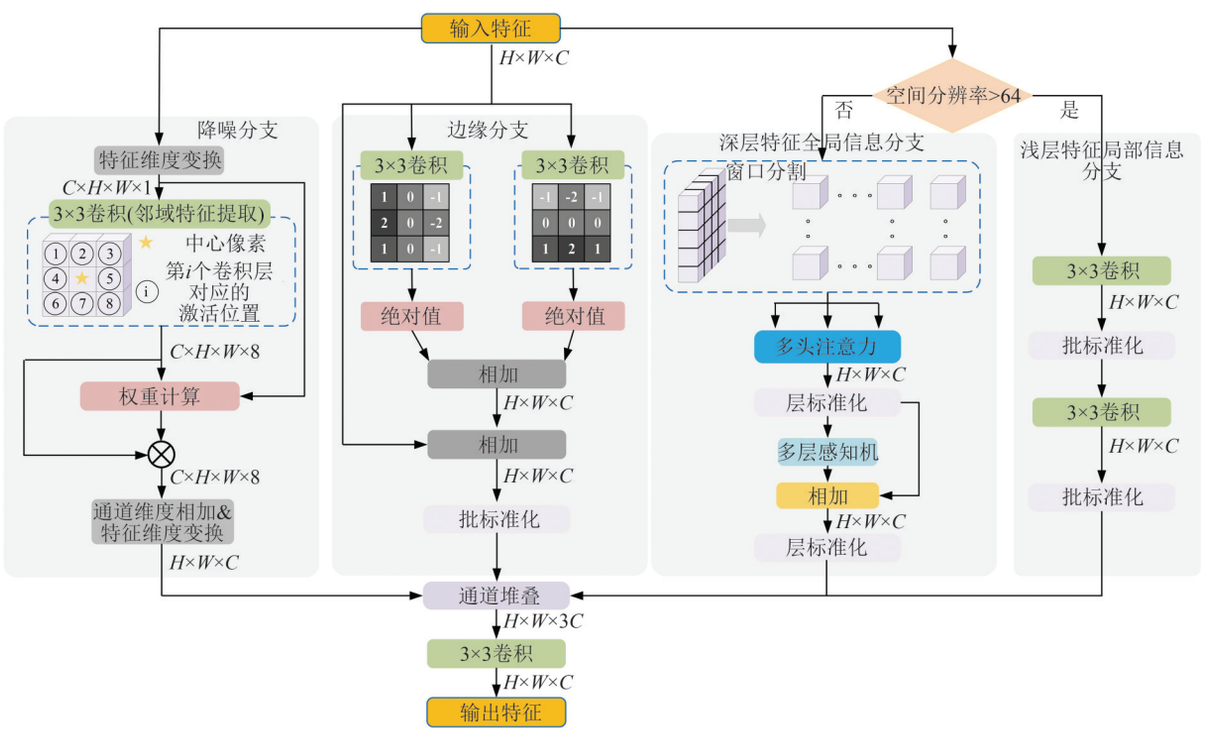


图3　结构边缘增强模块

Fig. 3　Structure-Aware Edge Enhancement Module

边缘分支用于为网络提供显式的结构边界先验。该分支采用水平和垂直方向的Sobel卷积核提取输入特征中的梯度响应,并将两个方向的边缘信息进行融合,获得地物轮廓和类别交界区域的结构提示。与完全依赖数据驱动方式学习边界不同,Sobel边缘响应能够在点标签极少的条件下为模型提供较为稳定的边界位置约束。随后,边缘响应与输入特征进行残差式融合,使网络在保留原始语义表达的同时增强边界、轮廓和细小结构区域的响应。

除降噪和边缘增强外,SAEE模块还根据输入特征的空间分辨率自适应选择不同的上下文建模方式。当输入特征空间分辨率较高时,浅层特征通常包含更多纹理、边界和小尺度形态信息,因此采用浅层特征局部信息分支,通过连续3×3卷积提取局部结构细节,以保留车辆、窄道路和建筑边缘等细粒度目标的空间信息。当输入特征空间分辨率较低时,深层特征具有更强的语义抽象能力,但局部卷积难以充分建模长距离依赖关系,因此模块采用深层特征空间信息分支。该分支先对特征进行窗口划分,再通过多头注意力和多层感知机建模窗口内的空间关系,并

结合残差连接保持原始语义信息,从而增强大范围地物区域的上下文一致性。

最后,SAEE模块将降噪分支、边缘分支以及尺度自适应上下文分支得到的特征在通道维度进行堆叠,完成特征融合与输出映射。通过上述设计,SAEE模块不仅利用显式边缘信息强化类别交界和目标轮廓,还通过邻域权重抑制不稳定响应,并根据不同尺度特征的表达特点自适应补充局部细节或空间上下文信息,从而提升点标签稀疏条件下模型对地物边界、细小目标和复杂空间结构的表达能力。

2 性能评估

2.1 数据集与对比方法

本文在ISPRS Potsdam和ISPRS Vaihingen两个高分辨率遥感语义分割数据集上进行实验,两个数据集均使用光学影像作为主源数据,并使用数据集配套的DSM作为辅助源数据,二者在空间上逐像素配准。所有影像在保持主、辅源空间对应关系的前提下裁剪为256×256图块。Potsdam数据集的训练集与验证集来自32幅原始图幅,按图块级划分后得到13 824个训练图块和4 608个验证图块,训练集与验证集共享同一原始图幅池,但具体坐标图块互斥;另选6幅独立图幅构建测试集,共得到3 456个测试图块。Vaihingen数据集的训练-验证数据来自11幅原始图幅,预处理后得到2 160个训练图块和256个验证图块;另选5幅独立图幅构建测试集,共得到412个测试图块。此外,训练集会按轮次随机打乱,验证集在每轮训练结束后进行一次评估。训练阶段并没有使用任何数据增强方法。以验证集像素总体精度作为最优模型选择依据,并保存验证精度最高的模型。

实验采用点注释弱监督设置,对于每幅256×256图像,仅随机保留0.01%的像素点标签作为监督信息,约对应6或7个有标签像素,其余像素在训练损失中被忽略。训练时仅对有点标签的位置计算监督损失,损失函数为交叉熵损失,且未标注像素不参与损失计算,从而模拟极低标注成本下的遥感语义分割场景。

实验基于PyTorch框架,在Windows操作系统和单个NVIDIA GeForce RTX 3060 GPU上完成,软件环境为Python 3.10和CUDA 12.2。输入为逐像素配准的3通道光学影像和1通道DSM,图块尺寸为256×256。主源编码器采用SAM ViT-B预训练权重初始化,并在训练过程中冻结其图像编码器参数;辅助源采用ImageNet预训练的ResNet18,并将首层卷积调整为单通道输入。模型采用Adam优化器,初始学习率为1×10^{-3},权重衰减为1×10^{-3},批量大小为6,共训练30轮。

为全面评价所提方法,本文从网络架构、监督形式和多源融合策略3个维度选择对比模型。HRNetv2[12]和UNetFormer[13]分别代表卷积多尺度建模与Transformer类遥感语义分割架构,用于检验通用分割骨干在极稀疏点监督下的表现;DBFNet[32]和AESAM[33]代表面向点监督或弱监督遥感分割的专门方法;CMFNet[25]和FTransUNet[12]代表多源或多模态特征融合方法;PAMSNet[20]是与本文任务设置最接近的点注释驱动多源遥感语义分割方法。所有对比方法均采用相同的数据划分、点注释掩膜、256×256输入尺寸和评价指标,以保证对比公平。

2.2 Potsdam数据集对比结果

表1给出了不同方法在Potsdam数据集上的定量对比结果。DGSE-Net的总体精度(overall accuracy,OA)为82.69%,平均交并比(mean intersection over union,mIoU)为54.07%,均高于其他对比方法。与PAMSNet相比,DGSE-Net的OA提高2.53个百分点,mIoU提高1.35个百分点,说明在相同点注释弱监督条件下,细节-全局融合与结构增强能够有效提升多源特征利用效率。从各类别交并比(intersection over union,IoU)可以看出,DGSE-Net在不透水面类别上表现较好,达到72.20%,说明该方法能够较稳定地识别大面积连续区域;低矮植被、树木和汽车的IoU分别为65.72%、55.72%和67.55%,也保持

表1 不同方法在Potsdam数据集上的精度对比/%

Table 1 Accuracy Comparison of Different Methods on the Potsdam Dataset/%

对比方法	总体指标		各类别IoU				
	OA	mIoU	不透水面	建筑	低矮植被	树木	汽车
HRNetv2	80.44	53.70	70.95	14.54	67.94	49.87	65.20
UNetFormer	6.32	40.39	48.86	0.69	47.49	41.02	63.90
DBFNet	79.05	50.47	69.26	10.08	62.26	46.98	63.75
AESAM	81.29	52.48	67.75	8.72	57.37	60.98	67.57
CMFNet	74.15	42.11	63.52	1.15	47.51	41.91	56.47
FTransUNet	75.63	43.99	61.48	8.65	40.21	48.21	61.41
PAMSNet	80.16	52.72	69.63	16.27	62.36	50.80	64.55
DGSE-Net	82.69	54.07	72.20	9.15	65.72	55.72	67.55

了较有竞争力的结果。需要指出的是,建筑类别IoU为9.15%,并非所有方法中的最高值,说明在Potsdam数据集的极稀疏点监督条件下,建筑区域仍存在明显混淆和边界恢复不足。

结合图4的可视化结果可以看出,DGSE-Net在植被边界、道路和细小车辆区域能够获得相对连续的预测,局部噪声少于部分对比方法;在复杂建筑密集区域,模型虽然能够保持一定轮廓结构,但建筑类别仍存在漏分和误分现象。这与表1中建筑类别IoU不占优的定量结果一致,也表明后续仍需进一步增强建筑边界和屋顶结构的弱监督约束。

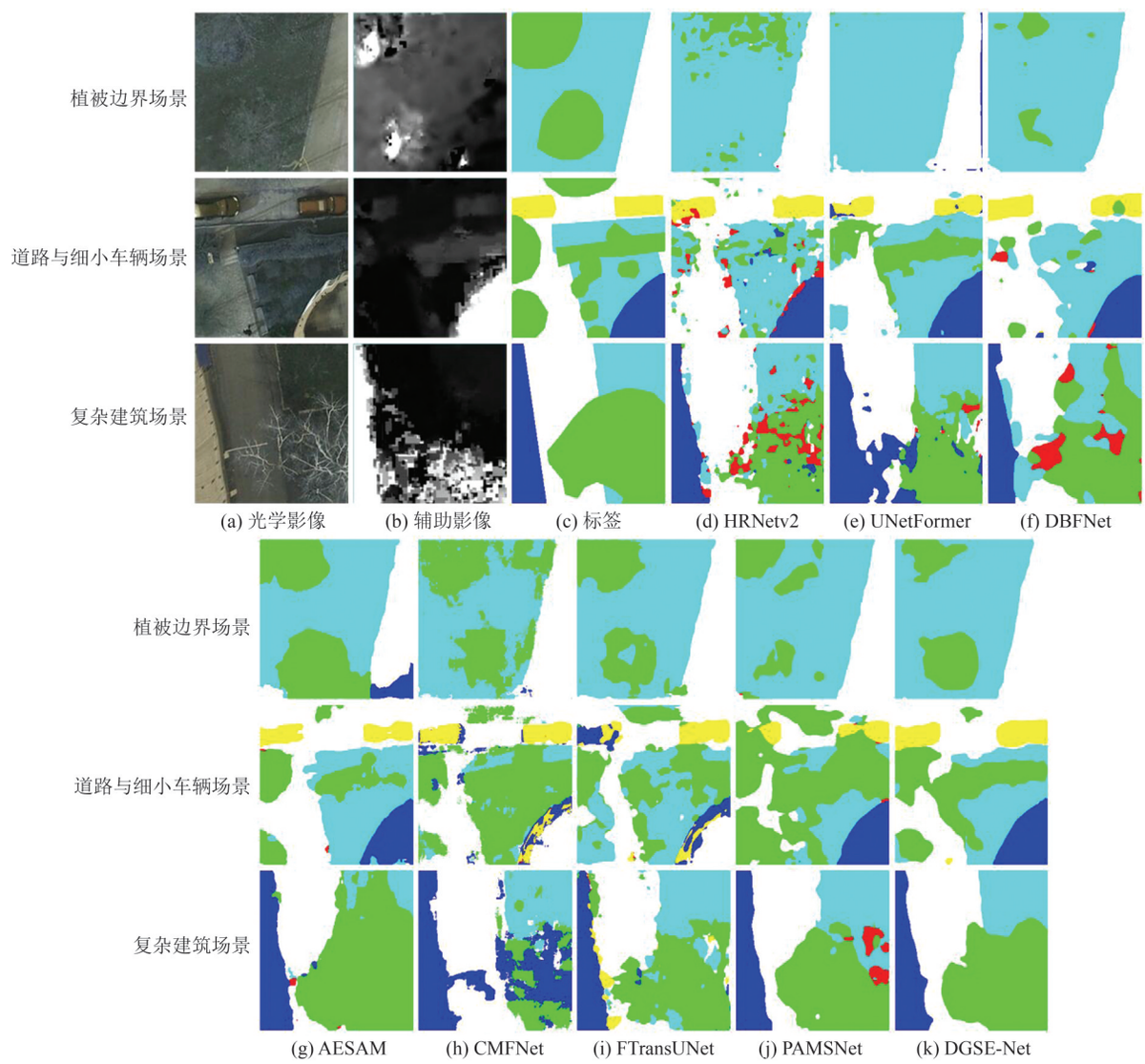


图4　Potsdam数据集可视化对比结果

Fig. 4　Visual Comparison Results on the Potsdam Dataset

2.3　Vaihingen数据集对比结果

表2列出了不同方法在Vaihingen数据集上的定量对比结果。DGSE-Net的OA为82.19%,mIoU为63.46%,均高于其他对比方法。与PAMSNet相比,OA提高1.73个百分点,mIoU提高2.46个百分点,表明所提方法在不同数据集上均能稳定提升多源弱监督分割性能。从各类别IoU结果来看,DGSE-Net在所有类别均取得了最优分割精度。其中建筑、不透水面和树木类别具有较高精度,说明SAEE对地物结构边界和层级上下文的增强有助于恢复具有明确轮廓或空间聚集特征的目标;汽车类别虽然仍难以获得较高精度,但相较多数对比方法具有更好表现,反映了细节增强分支对小目标有一定补充作用。

图5的可视化结果进一步表明,DGSE-Net在建筑边缘、树冠交界和小型车辆区域的预测更完整,局部碎片化现象有所减少。对于低矮植被与树木相邻区域,仍可能出现局部类别混淆,主要原因是点注释监督过于稀疏,难以为纹理相近类别提供充分区域约束。

表2 不同方法在Vaihingen数据集上的精度对比/%

Table 2 Accuracy Comparison of Different Methods on the Vaihingen Dataset/%

对比方法	总体指标		各类别IoU				
	OA	mIoU	不透水面	建筑	低矮植被	树木	汽车
HRNetv2	77.25	55.12	66.62	74.04	49.24	64.35	21.33
UNetFormer	78.98	56.68	69.20	76.57	50.15	66.25	21.24
DBFNet	76.58	54.63	65.76	73.87	48.83	63.05	21.65
AESAM	77.78	57.72	66.57	77.12	48.77	64.00	32.16
CMFNet	75.59	52.01	66.19	73.33	44.30	62.07	14.19
FTransUNet	78.66	58.60	66.90	79.83	47.94	66.58	31.77
PAMSNet	80.46	61.00	71.30	80.94	54.67	65.81	32.29
DGSE-Net	82.19	63.46	73.51	82.71	56.59	69.27	35.24

2.4 消融实验

表3给出了DGCF和SAEE的消融实验结果。基线模型的OA和mIoU分别为78.41%和56.46%。仅加入DGCF后,OA和mIoU分别提高2.42和2.81个百分点,说明细节-全局多源融合能够有效挖掘主源与辅助源之间的互补信息。仅加入SAEE模块后,OA和mIoU相比基线分别提高1.59和3.13个百分点,表明结构边缘增强对mIoU的提升更明显,尤其有助于改善边界和小目标类别的分割质量。例如汽车IoU由基线的19.58%提升至28.09%,说明显式边缘与局部结构约束能够缓解弱监督条件下细小目标监督不足的问题。

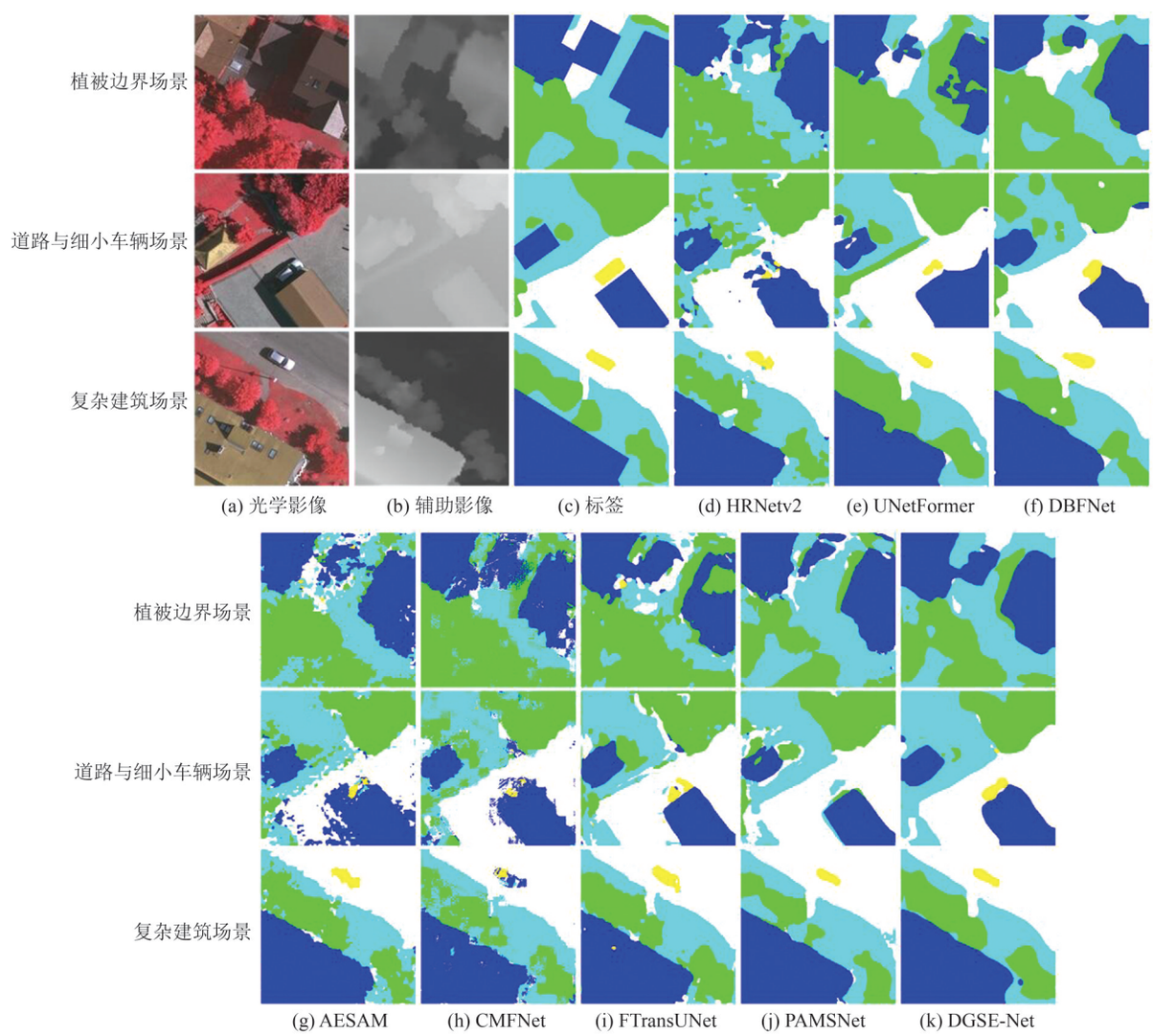


图5 Vaihingen数据集可视化对比结果

Fig. 5 Visual Comparison Results on the Vaihingen Dataset

同时加入DGCF和SAEE模块后,OA和mIoU分别达到82.19%和63.46%,为所有消融设置中的最优结果。与基线模型相比,OA提高3.78个百分点,mIoU提高7.00个百分点。DGCF模块更侧重于多源互补融合和全局语义交互,SAEE模块更侧重于边界、小目标和空间结构增强,二者联合能够形成互补作用,从而获得最佳分割性能。

表 3　DGCF 与 SAEE 模块消融实验/%

Table 3　Ablation Experiments of DGCF and SAEE Modules/%

DGCF	SAEE	总体指标		各类别 IoU				
		OA	mIoU	不透水面	建筑	低矮植被	树木	汽车
		78.41	56.46	67.60	79.28	52.81	63.05	19.58
√		80.83	59.27	74.47	80.97	54.83	64.70	21.38
	√	80.00	59.59	71.43	80.75	51.41	66.26	28.09
√	√	82.19	63.46	73.51	82.71	56.59	69.27	35.24

3　结　语

针对多源遥感影像在极稀疏点注释条件下语义分割精度受限的问题,本文提出了 DGSE-Net。该网络利用 SAM 图像编码器和 ResNet18 编码器分别提取主源与辅助源多尺度特征,通过 DGCF 在不同尺度上独立完成细节增强和全局交互,再借助 SAEE 强化边缘结构、小目标和上下文表达,最终由 CRFI 完成跨分辨率特征集成与语义预测。在 ISPRS Potsdam 和 ISPRS Vaihingen 数据集上的实验结果表明,DGSE-Net 在 0.01% 点注释弱监督设置下的 OA 分别达到 82.69% 和 82.19%,整体优于多种对比方法。消融实验进一步验证了 DGCF 和 SAEE 的有效性,DGCF 有助于提升多源互补融合能力,SAEE 有助于改善边界和小目标识别,二者联合时取得最佳性能。

目前仍存在需进一步研究的问题:一方面,SAM 编码器带来了较强视觉先验,但也可能增加模型复杂度,后续可探索更轻量的多源编码结构。另一方面,在跨区域、跨传感器场景中,点注释弱监督模型的泛化能力仍需加强。未来可进一步研究更少点标注、主动点选择或自动点标注策略,并结合更可靠的伪标签生成机制,以提高多源弱监督遥感语义分割的实用性。

参　考　文　献

[1] HUANG L, JIANG B, LV S, et al. Deep-Learning-Based Semantic Segmentation of Remote Sensing Images: A Survey[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 8370-8396.

[2] LI J, CAI Y, LI Q, et al. A Review of Remote Sensing Image Segmentation by Deep Learning Methods [J]. *International Journal of Digital Earth*, 2024, 17(1): 2328827.

[3] LI F T, ZHOU F D, ZHANG G, et al. HSAA-CD: A Hierarchical Semantic Aggregation Mechanism and Attention Module for Non-agricultural Change Detection in Cultivated Land[J]. *Remote Sensing*, 2024, 16(8): 1372.

[4] 冯权泷,陈泊安,李国庆,等. 遥感影像样本数据集研究综述[J]. 遥感学报, 2022, 26(4): 589-605.
FENG Quanlong, CHEN Boan, LI Guoqing, et al. A Review for Sample Datasets of Remote Sensing Imagery [J]. *National Remote Sensing Bulletin*, 2022, 26(4): 589-605.

[5] 吴志峰,曹峥,郑子豪,等. 中国城市遥感研究综述[J]. 遥感学报, 2025, 29(6): 2188-2215.
WU Zhifeng, CAO Zheng, ZHENG Zihao, et al. A Review of China Urban Remote Sensing[J]. *National Remote Sensing Bulletin*, 2025, 29(6): 2188-2215.

[5] 林娜,张小青,王岚,等. 基于空洞卷积U-Net的遥感影像道路提取方法[J]. 测绘地理信息,2025, 50(3):63-67.
LIN Na, ZHANG Xiaoqing, WANG Lan, et al. A Road Extraction Method for Remote Sensing Image Based on Dilated Convolutions U-Net[J]. *Journal of Geomatics*,2025,50(3):63-67.

[7] 王艳丽,董志鹏,王密. 基于双路卷积神经网络的高分辨率遥感影像浒苔检测方法[J]. 武汉大学学报(信息科学版), 2024, 49(12): 2261-2270.
LIN WANG Yanli, DONG Zhipeng, WANG Mi. Ulva Polifera Detection from High Resolution Remote Sensing Images Based on Dual-Path Convolutional Neural Networks[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(12): 2261-2270.

[8] 高常鑫,徐正泽,吴东岳,等. 深度学习实时语义分割综述[J]. 中国图象图形学报,2024,29(5): 1119-1145.
GAO Changxin, XU Zhengze, WU Dongyue, et al. Deep Learning-Based Real-Time Semantic Segmentation: A Survey [J]. *Journal of Image and Graphics*, 2024, 29(5): 1119-1145.

[9] DIMITROVSKI I, SPASEV V, LOSHKOVSKA S, et al. U-Net Ensemble for Enhanced Semantic Segmentation in Remote Sensing Imagery[J]. *Remote Sensing*, 2024, 16(12): 2077.

[10] CHANG Z, XU M, WEI Y, et al. UNeXt: An Efficient Network for the Semantic Segmentation of High-Resolution Remote Sensing Images[J]. *Sensors*, 2024, 24(20): 6655.

[11] WANG L, 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.

[12] MA X, ZHANG X, PUN M. O, et al. A Multilevel Multimodal Fusion Transformer for Remote Sensing Semantic Segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-15.

[13] HUO C, CHEN K, ZHANG S, et al. When Remote Sensing Meets Foundation Model: A Survey and Beyond [J]. *Remote Sensing*, 2025, 17(2): 179.

[14] MA X, WU Q, ZHAO X, et al. SAM-Assisted Remote Sensing Imagery Semantic Segmentation with Object and Boundary Constraints[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-16.

[15] LIU N, XU X, SU Y, et al. PointSAM: Pointly-Supervised Segment Anything Model for Remote Sensing Images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-15.

[16] 刘思涌,赵毅力.微调SAM的遥感图像高效语义分割模型DP-SAM[J].中国图象图形学报,2025,30(8):2884-2896.
LIU Siyong, ZHAO Yili. DP-SAM: Efficient Semantic Segmentation of Remote Sensing Images by Fine-Tuning SAM [J]. *Journal of Image and Graphics*, 2025, 30(8): 2884-2896.

[17] WANG S, CHEN W, XIE S. M, et al. Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery[J]. *Remote Sensing*, 2020, 12(2): 207.

[18] CHEN Z, LIAN Y, BAI J, et al. Weakly Supervised Semantic Segmentation of Remote Sensing Images Using Siamese Affinity Network[J]. *Remote Sensing*, 2025, 17(5): 808.

[19] 项伟康,周全,崔景程,等.基于深度学习的弱监督语义分割方法综述[J].中国图象图形学报,2024,29(5):1146-1168.
XIANG Weikang, ZHOU Quan, CUI Jingcheng, et al. Weakly Supervised Semantic Segmentation Based on Deep Learning[J]. *Journal of Image and Graphics*, 2024, 29(5): 1146-1168.

[20] ZHAO Y, JIA M, SUN G, et al. PAMSNet: A Point Annotation-Driven Multi-source Network for Remote Sensing Semantic Segmentation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 229: 1-16.

[21] 孙铭辰,葛洪伟,李婷.基于扩展补丁对的弱监督语义分割网络[J].信息与控制,2025,54(4):595-606.
SUN Mingchen, GE Hongwei, LI Ting. Weakly Supervised Semantic Segmentation Network Based on Extended Patch Pairs[J]. *Information and Control*, 2025, 54(4): 595-606.

[22] 李军侠,苏京峰,崔滢,等.基于语义调制的弱监督语义分割[J].软件学报,2025,36(9):4373-4387.
LI Junxia, SU Jingfeng, CUI Ying, et al. Semantic-Modulation-Based Weakly Supervised Semantic Segmentation[J]. *Journal of Software*, 2025, 36(9): 4373-4387.

[23] YU J, Zeng P, YU Y Y, et al. A Combined Convolutional Neural Network for UrbanLand-Use Classification with GIS Data[J]. *Remote Sensing*, 2022, 14: 1128.

[24] HONG D, HU J, YAO J, et al. Multimodal Remote Sensing Benchmark Datasets for Land Cover Classification with a Shared and Specific Feature Learning Model[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 178: 68-80.

[25] MA X, ZHANG X, PUN M O. A Crossmodal Multiscale Fusion Network for Semantic Segmentation of Remote Sensing Data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 3463-3474.

[26] 何友,刘瑜,谭大宁,等.多源遥感图像融合语义分割发展现状与展望[J].现代雷达,2024,46(2):16-24.
HE You, LIU Yu, TAN Daning, et al. Development Status and Prospect of Multi-source Remote Sensing Image Fusion Semantic Segmentation[J]. *Modern Radar*, 2024, 46(2): 16-24.

[27] 孙奥,徐芳,江树国,等.遥感基础模型多模态协同增强的语义分割方法[J].航空学报,2026,47(10):532910.
SUN Ao, XU Fang, JIANG Shuguo, et al. A Semantic Segmentation Method Enhanced by Multimodal Collaboration in Remote Sensing Foundation Models[J]. *Acta Aeronautica et Astronautica Sinica*, 2026, 47(10): 532910.

[28] 王定盼,董小环,黄令勇,等.信息熵估计辅助的域自适应多源遥感影像地表覆盖分类[J].武汉大学学报(信息科学版),2024,49(10):1940-1952.
WANG Dingpan, DONG Xiaohuan, HUANG Lingyong, et al. Information Entropy Uncertainty Estimation Based Domain Adaptation for Land Cover Classification from Multi-source Remote Sensing Images[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(10): 1940-

1952.

[29] 龚健雅,张展,贾浩巍,等. 面向多源数据地物提取的遥感知识感知与多尺度特征融合网络[J]. 武汉大学学报(信息科学版),2022,47(10):1546-1554.
GONG Jianya, ZHANG Zhan, JIA Haowei, et al. Multi-source Data Ground Object Extraction Based on Knowledge-Aware and Multi-scale Feature Fusion Network[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(10): 1546-1554.

[30] LIU C, SUN Y, XU Y, Et al. A Review of Optical and SAR Image Deep Feature Fusion in Semantic Segmentation[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17: 12910-12929.

[31] MA X, ZHANG X, PUN M O, et al. A Unified Framework with Multimodal Fine-Tuning for Remote Sensing Semantic Segmentation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-15.

[32] WU L, FANG L, YUE J, et al. Deep Bilateral Filtering Network for Point-Supervised Semantic Segmentation in Remote Sensing Images[J]. *IEEE Transactions on Image Processing*, 2022, 31: 7419-7434.

[33] CAO Y, HUANG X, WENG Q. A SAM-Adapted Weakly-Supervised Semantic Segmentation Method Constrained by Uncertainty and Transformation Consistency[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2025, 137: 104440.