

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Liu Erhu, Yuan Sijie, Li Haowen, Xu Shengjun, Hu Yu, Yang Tiantian. Multi-level feature extraction and edge-enhanced network for remote sensing building extraction[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260045. (刘二虎, 袁思杰, 李浩文, 徐胜军, 胡煜, 杨甜甜. 融合多级特征提取与边缘增强的遥感影像建筑物提取网络[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.260045.) [DOI: 10.11834/jig.260045]

融合多级特征提取与边缘增强的遥感影像建筑物提取网络

刘二虎¹, 袁思杰¹, 李浩文¹, 徐胜军¹, 胡煜², 杨甜甜³

1. 西安建筑科技大学信息与控制工程学院, 西安 710055; 2. 中科星图股份有限公司, 北京 101399; 3. 陕西学前师范学院信息工程学院, 西安 710100

摘要: 目的 建筑物提取是遥感影像解译中的一项关键任务, 针对现有遥感影像建筑物提取方法对多尺度特征建模能力不足以及对建筑物边缘轮廓刻画不充分的问题, 提出了一种融合多级特征提取与边缘增强的遥感影像建筑物提取网络(multi-level feature extraction and edge-enhanced network, MFEE-Net)。方法 该网络采用编码器-解码器架构, 首先构建了一种轻量化多尺度特征提取编码器, 以残差多分支卷积块(residual multi branch convolution block, ResMBC)为核心单元, 在保持卷积网络局部建模优势的同时, 通过并行多尺度分支显式增强网络对不同尺度建筑物的表征能力; 其次, 设计了层间特征融合模块(interlayer feature fusion module, IFFM), 联合建模空间信息与通道相关性, 实现编-解码器异质特征的自适应融合; 最后, 在解码阶段构建了一种边缘感知增强模块(edge-aware enhancement module, EAEM), 并结合边缘约束的联合损失函数, 对建筑物主体区域与边界细节进行联合优化, 从而提升分割结果的完整性与边缘精细度。结果 在WHU航空影像建筑物和Massachusetts建筑物数据集上与多种方法进行了比较, 实验结果表明, 所提出的网络MFEE-Net在WHU航空影像建筑物数据集上IoU、F1-score、precision和recall分别可达91.13%、95.36%、95.81%和94.92%, F1-score比次优方法提升0.36%; 在Massachusetts建筑物数据集上IoU、F1-score、precision和recall分别可达到75.46%、86.01%、87.84%和84.26%, F1-score比次优方法提升0.92%。结论 本文所提出的融合多级特征提取与边缘增强的遥感影像建筑物提取网络, 通过多尺度特征建模与边缘感知增强的协同作用, 可以有效提升遥感影像建筑物分割的整体精度和边界质量。

关键词: 遥感影像; 建筑物提取; 轻量化编码器; 残差多分支卷积; 层间特征融合; 边缘感知增强

Multi-level feature extraction and edge-enhanced network for remote sensing building extraction

Liu Erhu¹, Yuan Sijie¹, Li Haowen¹, Xu Shengjun¹, Hu Yu², Yang Tiantian³

1. School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China; 2. Geovis Technology Co., Ltd, Beijing 101399, China; 3. School of Information Engineering, Shaanxi Xueqian Normal University, Xi'an 710100, China

收稿日期: 2026-01-19; 修回日期: 2026-04-16

基金项目: 国家自然科学基金项目(62476211); 陕西省自然科学基金基础研究计划一般项目(2024JC-YBMS-483, 2025JC-YBQN-954, 2025JC-YBQN-857); 陕西省教育厅自然科学基金专项(24JK0513)

Supported by: National Natural Science Foundation of China (Grant No. 62476211); General Project of the Shaanxi Provincial Natural Science Foundation Research Program (2024JC-YBMS-483, 2025JC-YBQN-954, 2025JC-YBQN-857); Natural Science Special Research Program of the Shaanxi Provincial Department of Education (24JK0513)

Abstract: Objective Building extraction from remote sensing imagery is a fundamental and challenging task in remote sensing image interpretation and has been widely applied in urban planning, land-use analysis, disaster assessment, and geographic information system updating. With the rapid development of high-resolution and very-high-resolution (VHR) remote sensing sensors, buildings in aerial images exhibit large variations in scale, shape, texture, and spatial distribution. In addition, complex backgrounds, shadow interference, and spectral similarity between buildings and surrounding objects further increase the difficulty of accurate building extraction. Although convolutional neural network (CNN) based semantic segmentation methods have achieved remarkable progress in this field, existing approaches still face two major limitations. First, many networks lack sufficient capability to model multi-scale building features, resulting in missed detections or incomplete segmentation of buildings with large scale variations. Second, due to repeated down-sampling operations and insufficient boundary supervision, the extracted building boundaries are often blurred or discontinuous, which degrades the geometric accuracy of segmentation results. To address these issues, this paper proposes a novel remote sensing building extraction network that integrates multi-level feature extraction and edge enhancement, named the multi-level feature extraction and Edge-enhanced network (MFEE-Net), with the aim of improving both overall segmentation accuracy and boundary quality. **Method** The proposed MFEE-Net adopts an encoder-decoder architecture and is specifically designed to jointly enhance multi-scale feature representation and boundary detail preservation. In the encoding stage, a lightweight multi-scale feature extraction encoder is constructed using a newly designed residual multi branch convolution block (ResMBC) as the fundamental building unit. The ResMBC introduces parallel convolutional branches with different receptive fields, enabling the network to capture building structures and texture patterns at multiple spatial scales while retaining the local modeling advantages of standard convolution. The residual connection further facilitates stable training and effective feature propagation, allowing the encoder to generate rich and discriminative feature representations with relatively low computational cost. To effectively utilize features from different encoding depths, an interlayer feature fusion module (IFFM) is introduced between the encoder and decoder. Unlike simple skip connections, the IFFM jointly models spatial information and channel-wise correlations, enabling adaptive fusion of heterogeneous features from different layers. By enhancing feature complementarity and reducing semantic inconsistencies between low-level spatial details and high-level semantic representations, the IFFM alleviates information loss and improves the robustness of feature transmission during the decoding process. In the decoding stage, an edge-aware enhancement module (EAEM) is incorporated to explicitly refine building boundaries. The EAEM emphasizes edge-related features by enhancing boundary-sensitive responses and suppressing background interference, thereby improving the continuity and clarity of extracted building contours. Furthermore, a joint loss function with an edge-constrained auxiliary term is employed during training. This loss formulation encourages the network to simultaneously optimize building foreground regions and boundary details, leading to a coordinated improvement in segmentation completeness and edge precision. **Result** Extensive experiments were conducted on two widely used public benchmark datasets, namely the WHU Aerial Building Dataset and the Massachusetts Building Dataset, to evaluate the performance of the proposed MFEE-Net. Quantitative comparisons were performed against multiple state-of-the-art building extraction methods using standard evaluation metrics, including Intersection over Union (IoU), F1-score, precision, and recall. On the WHU Aerial Building Dataset, MFEE-Net achieved an IoU of 91.13%, an F1-score of 95.36%, a precision of 95.81%, and a recall of 94.92%, demonstrating superior performance in both overall accuracy and boundary consistency. On the Massachusetts Building Dataset, which contains lower-resolution imagery and poses greater challenges due to blurred edges and complex backgrounds, MFEE-Net attained an IoU of 75.46%, an F1-score of 86.01%, a precision of 87.84%, and a recall of 84.26%. These results indicate that the proposed network maintains robust performance under different spatial resolutions and scene complexities. Qualitative visual comparisons further reveal that MFEE-Net is capable of producing more complete building regions with clearer and more continuous boundaries, particularly in scenes with dense buildings, complex structures, and significant scale variations. Ablation studies validate the effectiveness of each proposed component, confirming that multi-level feature extraction, interlayer feature fusion, and edge-aware enhancement collaboratively contribute to performance improvements. **Conclusion** This study proposes a novel remote sensing building extraction network, MFEE-Net, which integrates multi-level feature extraction and edge enhancement within an encoder-decoder framework. By leveraging a lightweight multi-scale encoder, an interlayer feature

fusion strategy, and an edge-aware enhancement mechanism with boundary supervision, the proposed network effectively addresses the challenges of scale variation and boundary ambiguity in remote sensing building extraction. Experimental results on public benchmark datasets demonstrate that MFEE-Net achieves competitive and stable performance, significantly improving both segmentation accuracy and boundary quality. The proposed approach provides an effective solution for high-precision building extraction in complex remote sensing scenarios and offers potential for practical applications in urban analysis and geospatial information processing.

Key words: remote sensing images; building extraction; lightweight encoder; residual multi branch convolution; inter-layer feature fusion; edge-aware enhancement

0 引言

遥感技术通过搭载于卫星或航空平台的传感器,以非接触的方式获取大范围地表影像数据,为地表信息的获取与分析提供了重要支撑。遥感建筑物提取作为遥感影像解译中的一项关键任务,旨在实现对建筑物区域的像素级精确识别,在城市规划、人口评估和环境监测等领域具有广泛的应用价值(Hosseinpour 等, 2022; Li 等, 2021; 戎誉 等, 2026)。随着高分辨率(very-high-resolution, VHR)遥感影像获取能力的不断提升,建筑物的结构与细节特征得以更加充分地呈现。然而,由于建筑物在形态、尺度及材质等方面具有显著的多样性,其在遥感影像中往往表现出复杂的光谱与纹理特征,这为建筑物的精确提取带来了较大挑战(余帅和汪西莉, 2021; Zhao 等, 2020)。

在深度学习技术兴起之前,遥感语义分割研究主要依赖基于人工特征的传统方法。此类方法通常通过人为设定的规则或算子,从遥感影像中提取能够表征地物属性的特征信息,例如几何形态、纹理分布、阴影关系以及边缘结构等(Dornaika 等, 2016; Sun 等, 2015; Xie 等, 2024; Zhu 等, 2024)。随后,将这些特征输入支持向量机(support vector machine, SVM)、AdaBoost、随机森林等经典机器学习分类器,或结合条件随机场等概率图模型,对影像中的像素或图像单元进行类别判定,从而完成语义分割任务。Turker 和 Koc-San (2015)通过 SVM 对像素和区域进行分类以初步定位建筑候选区域,随后结合 Hough 变换提取直线结构,并引入感知分组(perceptual grouping)的几何约束,对建筑物的规则形状进行组织与优化,从而实现建筑物提取。Zhu 等人(2020)通过引入多尺度感知特征和分割先验约束(segmentation prior),在条件随机场(conditional random field, CRF)框架下对建筑物提取结果进行全局优化,有效提升了高分辨率遥感影像中建筑物区域的空间一致性与边界完整性。然而,由于传统方法的特征设计高度依赖人工经验,且难以同时刻画复杂场景中的多尺度信息与高层语义表达,这类方法在复杂地物分布和高分辨率遥感影像条件下往往表现出一定的局限性。

近年来,随着深度学习技术的快速发展,基于卷积神经网络(convolutional neural network, CNN)的方法在图像语义分割领域取得了显著进展,并逐渐成为遥感建筑物提取研究的主流技术路线(Feng 等, 2023; Varshosaz 等, 2024; Xie 等, 2020; 李国燕 等, 2023)。相比传统依赖人工设计特征的方法, CNN 能够通过多层卷积结构自动学习遥感影像中的多层次特征表示,涵盖纹理模式、光谱信息以及空间上下文关系,从而有效提升对复杂地物目标的判别能力。基于 CNN, Long 等人(2015)提出的全卷积网络(fully convolutional network, FCN)实现了端到端的像素级分类,为语义分割任务提供了统一的网络框架,并被广泛应用于高分辨率遥感影像的建筑物提取研究中。在 FCN 框架的基础上,研究者围绕网络结构设计及上下文信息建模对语义分割方法进行了大量改进。其中, U-Net 系列方法通过编码器-解码器结构及跳跃连接机制,将高层语义信息与低层空间细节有效融合,显著提升了分割结果的定位精度(Ronneberger 等, 2015; 金澍 等, 2023)。DeepLab 系列方法则通过引入空洞卷积(atrous convolution)和空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块,在不降低特征图分辨率的情况下扩大网络感受野,在复杂场景下表现出良好的分割性能(Chen 等, 2017)。针对高分辨率遥感影像中建筑物尺度差异大、结构复杂的问题,研究者进一步引入多

© 中国图象图形学报版权所有

尺度特征建模与融合策略。DeepResUnet将ResNet引入U-Net网络的编码器部分,实现了对建筑物更深层次特征的提取,增强了网络对建筑物局部细节特征的捕获能力(Onojeghuo等,2023)。CFF-Net(cross-hierarchy feature fusion network)在解码阶段末尾引入预测增强模块,该模块可同时对中间特征与最终输出特征进行预测,用于细化网络预测结果与真实标签之间的残差,从而进一步提升建筑物提取精度(Qian等,2024)。MFFNet(multi-level feature fusion network)通过多路径卷积结构融合不同层次和尺度的特征信息,并结合大感受野卷积与金字塔池化模块增强全局语义表达能力,从而改善建筑物提取的完整性与边界表现(李星华等,2022)。此外,EGAFNet(edge guidance adaptive fusion network)通过引入边缘引导模块和尺度感知自适应融合模块,在编码与解码过程中显式强化边界信息表达,并自适应地聚合多尺度特征,从而提升建筑物边缘轮廓的清晰度和尺度适应能力(Yang等,2025)。尽管上述方法在多尺度特征建模或边缘信息利用方面取得了积极进展,但其特征融合过程多侧重于特征的直接叠加或加权聚合,对编解码阶段异质特征之间的空间关系和通道相关性建模仍相对有限。

总体来看,基于卷积神经网络的语义分割方法可以在保持图像完整性的前提下通过特征融合、预测增强及注意力机制等,在局部空间细节和几何结构表达方面展现出良好的优势。为进一步增强神经网络的全局语义信息建模能力,近年来,基于长程依赖建模的Transformer模型被引入计算机视觉领域(Chen等,2021;Dosovitskiy,2020;Li等,2023),如DPENet(Chen等,2023)、DSFENet(Xia等,2023)、BuildFormer(Wang等,2022)等方法采用Transformer和CNN双路径进行全局与局部信息的提取,最终实现遥感图像的建筑物提取。然而,Transformer类方法通常模型复杂,计算开销大,在样本数量有限的情况下也更容易产生过拟合问题(Chen等,2022;Fu等,2024;Hu等,2023;Pan等,2024)。Mamba由于在建模长距离依赖关系的同时能够保持线性计算复杂度,近年来被视为Transformer的一种潜在替代方案(Ma等,2025;Wang等,2025;Zhao等,2025)。现阶段,尽管将CNN与Transformer或者Mamba结合的方法可以同时兼顾遥感影像的全局和局部特征,但基于图像Patch的处理方式往往会破坏影像空间几

何结构的连续性,容易导致小尺度建筑或细长结构(如屋檐、狭长建筑)信息的丢失,并引起目标边界模糊。此外,CNN提取的局部特征与基于Patch的全局特征在表征形式上的异质性,也可能引发语义差异,为多尺度特征的有效融合带来挑战。

基于此,本文针对现有网络难以从遥感影像中捕获建筑物的多尺度特征信息以及边缘轮廓模糊的问题,提出了一种融合多级特征提取与边缘增强的遥感影像建筑物提取网络(multi-level feature extraction and edge-enhanced network, MFEE-Net)。该网络采用编码器-解码器架构,首先设计了一种轻量化的多尺度特征提取主干网络;其次在编码器和解码器之间设计了层间特征融合模块(interlayer feature fusion module, IFFM),有效融合编解码过程中的异质特征;最后引入了边缘感知增强模块(edge-aware enhancement module, EAEM)以细化建筑物的边缘轮廓信息,并采用联合损失函数对网络进行训练,以优化建筑物分割的整体效果和边缘细节。所提网络贡献如下:

- 1)构建了一种轻量化多尺度编码器网络,以所设计的残差多分支卷积块(residual multi branch convolution block, ResMBC)为核心单元,在保持卷积网络局部建模优势的同时引入并行多尺度特征提取机制,提升了编码阶段对建筑物结构与纹理信息的综合表征能力。

- 2)设计了一种层间特征融合模块IFFM,联合建模空间信息与通道相关性,有效提升了不同尺度特征之间的互补性,减少了信息丢失与语义不匹配。

- 3)设计了一种边缘感知增强模块EAEM,通过在解码阶段强化边缘相关特征表达,并在训练过程中引入边缘约束的辅助损失,实现了对建筑物主体区域与边界细节的协同优化,并改善分割结果中边界模糊和轮廓不连续的问题。

- 4)提出了一种融合多级特征提取与边缘增强的遥感影像建筑物提取网络MFEE-Net,在公开的WHU和Massachusetts遥感建筑物语义分割数据集上进行了对比实验,在多个指标上均取得了优异的性能,验证了所提算法的有效性。

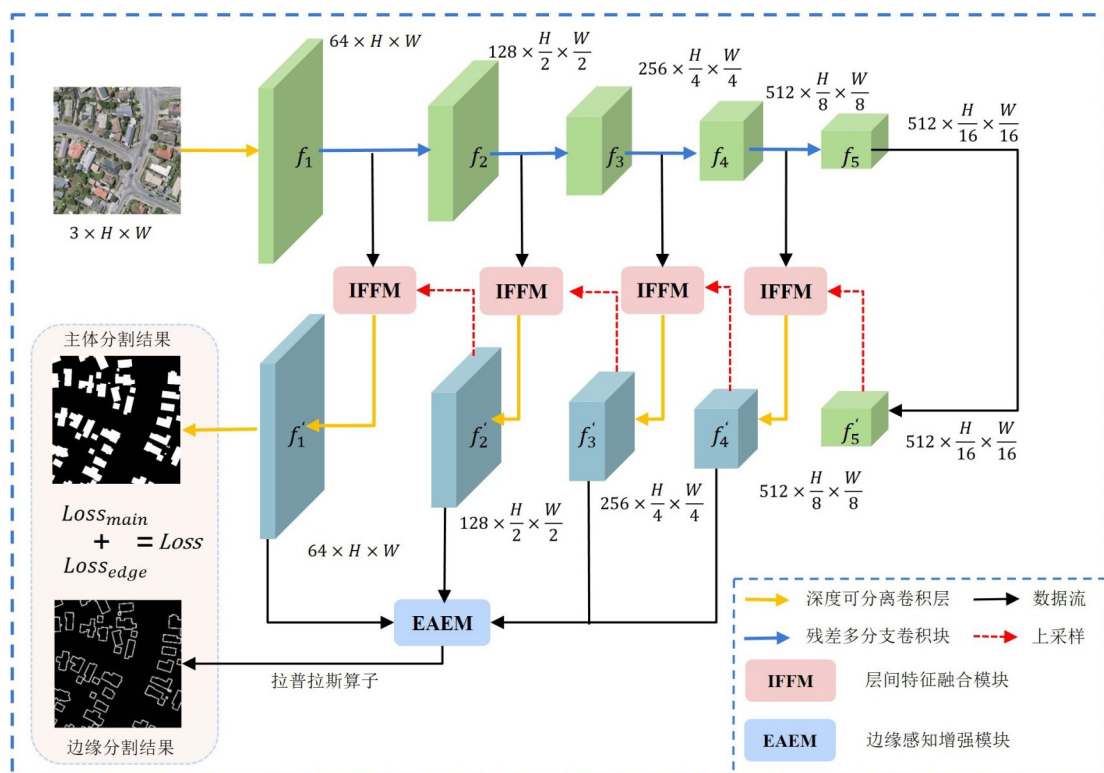


图1 网络整体结构图

Fig. 1 Overall structure of the network

1 研究方法

1.1 网络整体架构

本文所提的融合多级特征提取与边缘增强的遥感影像建筑物提取网络 MFEE-Net 的整体结构如图 1 所示。该网络采用基于 U-Net 的编码器 - 解码器结构。编码器由本文设计的 ResMBC 模块构成, 其通过不同大小的卷积分支提取遥感影像中的多尺度信息。解码器部分采用深度可分离卷积以减少网络的参数量, 并且引入层间特征融合模块 IFFM 对编码器和解码器获取的不同特征进行充分融合, 增强特征间的相关性。最后, 通过引入边缘感知增强模块 EAEM 对不同尺度的特征所包含的建筑物边缘信息进行增强, 以提高网络对于遥感影像中建筑物边缘轮廓的提取精度。

1.2 编码器和解码器

为在有效处理高分辨率遥感影像中建筑物差异性的同时兼顾计算效率, 本文在网络的编码器阶段设计了一种轻量级的特征提取主干网络, 主要由基本单元 ResMBC Block 构成, 具体结构如图 2 所示。

从结构上看, ResMBC Block 由两部分组成: 第一部分为深度可分离卷积层, 第二部分为多支路并行卷积 (multi branch parallel convolution)。其中, 深度可分离卷积层用于提取浅层特征并改变网络的通道数; 多支路并行卷积则通过不同尺寸的卷积层 (包括 3×3 卷积、 1×11 卷积、 11×1 及 1×1 卷积) 提取多尺度特征, 在增强网络的感受野的同时降低了网络参数量。其中, $DW1 \times 11$ 卷积与 $DW11 \times 1$ 卷积在感受野建模方向上具有明显差异: $DW1 \times 11$ 卷积主要用于建模水平方向的长距离上下文依赖, 而 $DW11 \times 1$ 卷积则侧重于捕获垂直方向的结构信息。二者的组合能够以较低的计算开销实现近似于大尺度二维卷积的效果, 在保持方向敏感性的同时增强网络对建筑物长条状结构与边缘轮廓的表达能力。

在整体架构设计上, 特征提取网络采用分阶段的层级结构形式, 可划分为两个主要阶段。第一阶段通过深度可分离卷积模块对输入遥感影像进行初步特征编码, 重点提取低层次的纹理与边缘信息。该模块由两个结构相同的深度可分离卷积层串联构成, 每个卷积层均采用大小为 3×3 、填充为 1 的卷积核, 并依次连接批量归一化层 (batch normalization,

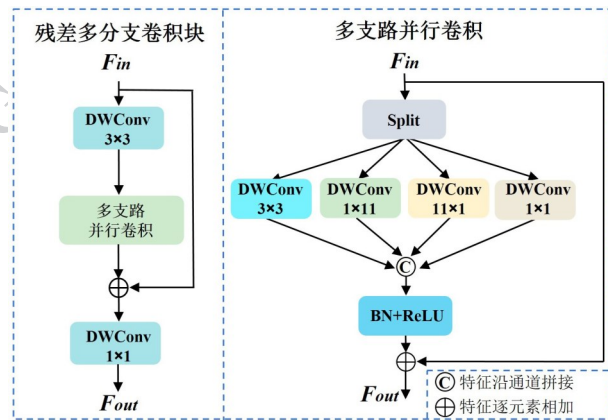


图2 ResMBC Block 结构图

Fig. 2 Structure of the ResMBC Block

BN)与线性整流激活函数(rectified linear unit, ReLU),以增强特征表达能力并加快网络收敛。第二阶段以 ResMBC Block 为基本单元构建四层特征提取结构,其中每层结构包含一个用于下采样操作的核大小为 2×2 的最大池化层及两组 ResMBC Block。总体来讲,编码器的特征提取网络对遥感影像进行多级特征编码,在逐步压缩空间分辨率的同时增强语义表达能力。编码器结构具体配置如表 1 所示。

在解码阶段,本文采用了一种渐进式上采样策略,通过逐层重建特征图分辨率,实现对像素级语义信息的精细表达。解码器整体由 5 个阶段构成,包含 5 个上采样模块、4 个层间特征融合模块 IFFM 以及 1 个边缘感知增强模块 EAEM。其中,前四个阶段均由一个上采样模块与一个 IFFM 模块组成,分别用于逐步提升特征图的空间分辨率,并融合来自编码器与解码器不同层级的多尺度语义信息。IFFM 模块通过整合浅层空间细节与深层语义特征,增强了

特征表示的完整性与一致性。在具体实现上,每个上采样模块采用双线性插值对特征图进行尺度恢复,随后通过一个深度可分离卷积层对上采样特征进行细化。该深度可分离卷积层由两个卷积核大小为 3、步长为 1 的深度可分离卷积构成,并在每个卷积后依次连接批量归一化层 BN 和线性整流激活函数 ReLU,以增强特征表达能力并稳定网络训练过程。在解码器的最后一个阶段,引入 EAEM 模块对融合后的多级特征进行边缘信息强化,并结合双线性插值完成最终的分辨率恢复。该设计在有效抑制上采样过程中棋盘效应的同时,增强了网络对建筑物边界与结构细节的表达能力。

1.3 层间特征融合模块 IFFM

为有效融合编码器与解码器中不同层级、不同尺度的特征信息,并增强相邻特征之间的相关性,本文设计了用于编解码器层间特征融合的 IFFM 模块。该模块旨在特征融合过程中同时建模空间信息与通道维度的语义关联,从而实现多尺度特征的自适应协调与互补表达,减少由于尺度差异和语义不一致带来的信息损失。IFFM 的整体结构如图 3 所示,主要由空间增强、特征关联以及空间映射归一化三个阶段组成。在空间增强阶段,模块通过对来自编码器与解码器的特征进行空间维度上的融合,突出具有判别性的空间区域,为后续特征交互提供更具上下文感知能力的融合特征。随后,在特征关联阶段,通过在通道维度上建模不同特征之间的相关性,引导网络自适应地强调互补信息并抑制冗余特征,从而提升特征表达的一致性。最后,在空间映射归一化阶段,对融合后的特征进行统一映射与归一化处理,以稳定特征分布并进一步增强跨层特征的协

表 1 编码器的结构配置

Table 1 Structure configuration of encoder

网络层级	输入通道	输出通道	池化层	特征图大小
输入层	3	-	-	$H \times W$
DWConv layer $\times 2$	3	64	-	$H \times W$
ResMBC Block $\times 2$	64	128	最大池化(2×2)	$H/2 \times W/2$
ResMBC Block $\times 2$	128	256	最大池化(2×2)	$H/4 \times W/4$
ResMBC Block $\times 2$	256	512	最大池化(2×2)	$H/8 \times W/8$
ResMBC Block $\times 2$	512	512	最大池化(2×2)	$H/16 \times W/16$

注:表中“-”表示不存在或者不适用。

同表达能力。

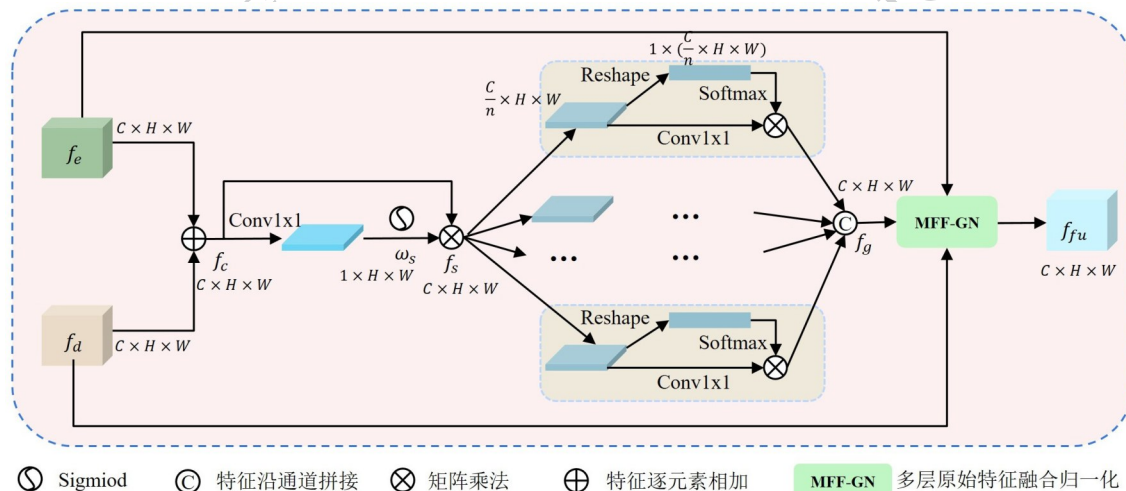


图3 IFFM结构图

Fig. 3 Structure of the IFFM

在空间增强部分,针对输入的两组特征 $f_e \in \mathbf{R}^{C \times H \times W}$ 及 $f_d \in \mathbf{R}^{C \times H \times W}$,首先对其在空间维度进行逐像素相加,以获取到融合后的粗略特征 $f_c \in \mathbf{R}^{C \times H \times W}$,为了细化该特征并获取上下文感知信息,使用 1×1 的卷积将特征压缩到单个通道中以聚合空间信息,然后通过 Sigmoid 激活函数生成空间注意力权重 ω_s 。得到空间信息加权后的特征 $f_s \in \mathbf{R}^{C \times H \times W}$,这一过程可以用公式(1)和(2)表示为:

$$\omega_s = \sigma(\text{Conv}(f_c \oplus f_d)) \quad (1)$$

$$f_s = f_c \otimes \omega_s \quad (2)$$

式中, $\sigma(\cdot)$ 表示 Sigmoid 函数, $\text{Conv}(\cdot)$ 表示卷积核大小为 1×1 的卷积层, \oplus 表示元素求和, \otimes 表示矩阵乘法。

在特征关联部分,为了增强不同特征之间的相关性,将空间增强部分所融合后的特征 f_s 沿通道维度分为 n 组,并且在每一组的基础上进行特征交互,具体来说,使用卷积核大小为 1×1 的卷积操作来细化每一组特征 $[f_{s_i}^k]_{n=i} \in \mathbf{R}^{(C/n) \times H \times W}$ 相邻通道间的特征信息,每一组特征 $[f_{s_i}^k]_{n=i}$ 在通道维度进行增强,通过 Softmax 函数在通道维度上对注意力权重进行归一化以生成通道间特征相关性的注意力权重 ω_g ,然后将 ω_g 应用于细化的特征。最后,将每组特征在通道维度进行拼接,形成融合后的特征 f_g 。整个计算过程如公式(3)和(4)所示:

$$f_{g_i} = (\text{Softmax}(\text{F}([f_{s_i}^k]_{n=i}))) \otimes (\text{Conv}([f_{s_i}^k]_{n=i})) \quad (3)$$

$$f_g = [f_{g1}, f_{g2}, \dots, f_{g_i}] \quad (4)$$

式中, $\text{F}(\cdot)$ 和 $\text{Conv}(\cdot)$ 分别表示 Reshape 操作及卷积核大小为 1×1 的卷积层。 f_{g_i} 代表每一组通道交互后的特征。 $[\]$ 表示特征拼接 (concat) 操作。

在空间映射归一化部分,将通道维度加权的分组聚合特征 f_g 与原始的两组特征 f_c 和 f_d 输入至多层原始特征融合归一化层 (normalization layer with multilayered original feature fusion, MFF-GN) 中,对特征 f_g 的均值和标准差进行归一化,从而得到具有强相关性的融合特征 f_{fu} ,上述过程可以表示为公式(5):

$$f_{fu} = \frac{f_g - \text{mean}(f_c \oplus f_d)}{\text{std}(f_c \oplus f_d)} \quad (5)$$

式中, $\text{mean}(\cdot)$ 和 $\text{std}(\cdot)$ 分别表示均值和标准差。

从上述过程可知,IFFM 通过充分利用不同层级特征之间的语义信息,并在通道维度上建模特征的相关性,从而增强整体特征的表达。

1.4 边缘感知增强模块 EAEM

为进一步强化建筑物提取结果边界的完整性,提高轮廓提取精度,在解码器阶段设计了一种边缘感知增强模块 EAEM。该模块以解码端经层间特征融合模块 IFFM 获得的多级特征为输入,在充分利用其高层语义信息的基础上,重点增强与建筑物边缘相关的特征表达,从而提升网络对建筑物边界与结构细节的感知能力。考虑到仅依赖于采样恢复分辨率往往难以充分保留和突出边缘细节,EAEM 通过显式建模边缘特征并在训练过程中引入边缘约束的

辅助损失函数,对解码阶段的特征学习进行引导,实现对建筑物主体区域与边界信息的协同优化。

EAEM的具体结构如图4所示, f'_1, f'_2, f'_3, f'_4 为解码器阶段通过层间特征融合模块IFFM获取到的不同尺度的特征,这些特征包含了丰富的边缘纹理信息及高级语义信息,首先对这些特征通过 1×1 的卷积将通道数调整至64维;其次,使用不同采样率大小的双线性插值法将 f'_2, f'_3, f'_4 上采样至与 f'_1 同大小;

然后,沿通道维度对四个特征进行拼接,可得到融合后的特征 f'_c 。以上计算过程可以表示为公式(6):

$$f'_c = [\text{Conv}(f'_1), \text{Conv}(\text{up}(f'_2)), \text{Conv}(\text{up}(f'_3)), \text{Conv}(\text{up}(f'_4))] \quad (6)$$

式中, $\text{Conv}(\cdot)$ 表示卷积核大小为 1×1 的卷积层。 $\text{up}(\cdot)$ 代表上采样过程。

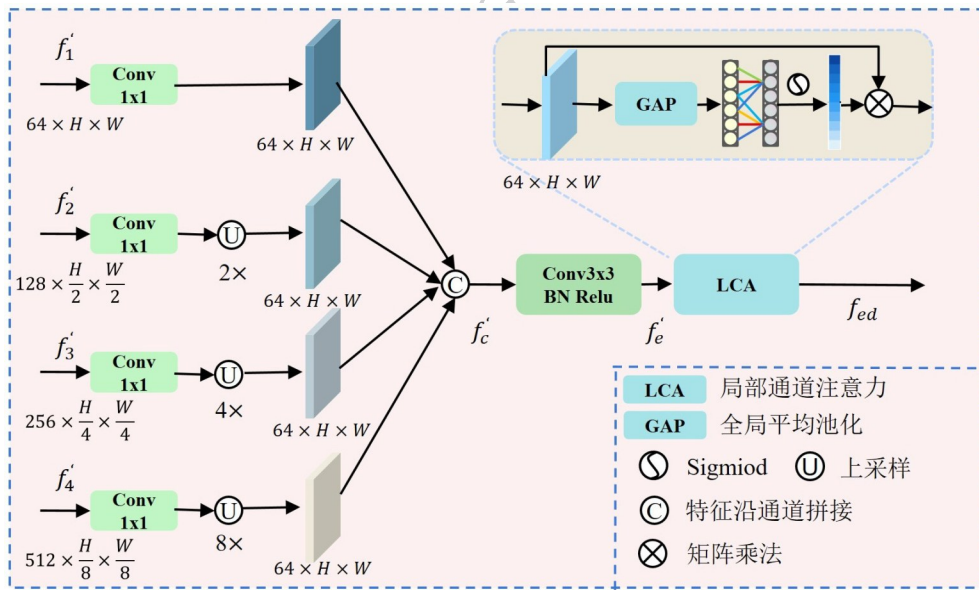


图4 EAEM结构图

Fig. 4 Structure of the EAEM

对于融合后的特征 f'_c ,首先通过一个 3×3 的卷积层进行通道压缩,得到边缘信息更加集中的特征 f'_e 。为进一步增强 f'_e 的边缘表达能力并挖掘其在通道维度上的语义关联,本文引入局部通道注意力机制对关键特征通道进行建模。具体而言,首先对特征 f'_e 进行全局平均池化以获得通道级特征描述,随后采用自适应卷积核大小的一维卷积实现跨通道信息交互,并通过Sigmoid激活函数生成通道权重系数。最后,将该权重系数与原始特征 f'_e 进行逐通道相乘,得到增强后的边缘特征 f_{ed} 。上述过程可用公式(7)表示为:

$$f_{ed} = \sigma(\text{Conv 1D}(\text{Avgpool}(f'_e))) \otimes f'_e \quad (7)$$

式中,Avgpool表示全局平均池化,Conv 1D表示自适应卷积核大小的一维卷积。从上述过程可知,所提出的EAEM模块可以有效融合高级语义信息与低级边缘信息,并通过注意力机制突出与边缘相关的关

键特征通道,抑制冗余通道及噪声干扰,从而增强边缘特征的语义表达能力。

1.5 联合损失函数

为实现建筑物的整体分割结果与边缘细节的同时优化,设计了联合损失函数进行监督学习。该损失函数由主分割损失与边缘约束的辅助损失共同构成,其中主损失用于监督建筑物主体的分割结果,辅助损失则通过引入边缘监督,引导网络更加关注建筑物边界区域的精细表达。主损失函数 \mathcal{L}_m 由交叉熵损失函数 \mathcal{L}_c 及骰子损失 \mathcal{L}_d 组成,其中交叉熵损失侧重于像素级分类精度,而骰子损失则主要缓解类别不平衡对模型训练带来的影响,提升分割完整度,具体可由公式(8)、(9)和(10)表示为:

$$\mathcal{L}_c = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^{(n)} \log \hat{y}_k^{(n)} \quad (8)$$

$$\mathcal{L}_d = -\frac{2}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^{(n)} \frac{\hat{y}_k^{(n)} y_k^{(n)}}{\hat{y}_k^{(n)} + y_k^{(n)}} \quad (9)$$

$$\mathcal{L}_m = \mathcal{L}_c + \mathcal{L}_d \quad (10)$$

式中, N 和 K 分别代表样本数量和类别数量, $y_k^{(n)}$ 代表预测标签, $\hat{y}_k^{(n)}$ 代表真实标签。

辅助损失函数 \mathcal{L}_a 首先使用拉普拉斯卷积核提取边缘感知增强模块 EAEM 输出特征的边缘像素及真实标签的边缘像素, 然后计算两者的二元交叉熵, 辅助损失函数如公式(11)所示:

$$\mathcal{L}_a = \mathcal{L}_b(\ell(Y), \ell(\hat{Y})) \quad (11)$$

式中, Y 代表 EAEM 模块的输出, \hat{Y} 代表真实标签, ℓ 代表用于提取预测标签和真实标签边界的拉普拉斯卷积核, \mathcal{L}_b 表示二元交叉熵函数。

为了更好地平衡主损失函数与辅助损失函数在训练过程中的贡献, 对辅助损失函数项引入权重系数 α 调整其对整个损失函数的占比, 因此最终的损失函数 \mathcal{L} 可以用公式(12)表示为:

$$\mathcal{L} = \mathcal{L}_m + \alpha \mathcal{L}_a \quad (12)$$

2 实验结果与分析

2.1 实验环境及参数设置

实验在 Ubuntu 20.04 操作系统环境下进行。实验平台配备 Intel Xeon Gold 6330 处理器、64 GB 内存, 并搭载一块 NVIDIA GeForce RTX 3090 显卡 (24 GB 显存)。网络模型的训练与测试基于 Python 3.8 和 PyTorch 1.10.0 深度学习框架实现, 同时采用 CUDA 11.3 与 cuDNN 8.0 作为 GPU 加速库, 以提升模型的训练与推理效率。在 MFEE-Net 网络的训练过程中, 为增强模型的泛化能力, 我们用随机翻转与归一化技术对划分后的训练集数据进行数据增强。训练的优化器选择 AdamW, 采用带重启机制的余弦退火学习率调整策略以动态更新学习率, 其中的参数设置如下: 初始周期长度 T_0 设置为 15, 周期倍增系数 T_{multi} 设置为 2, 初始学习率 η_0 设置为 1×10^{-3} , 最小学习率 η_{min} 设置为 1×10^{-6} 。训练的周期为 100 个 epoch, 批次大小为 8。

2.2 实验数据集及评价指标

实验使用 WHU 航空影像建筑物和 Massachusetts 建筑物两个公开遥感数据集。WHU 航空影像建筑物数据集的原始影像数据来自于新西兰土地信息服务网站, 经过数据集作者下采样后的地面分辨率为 0.3m, 影像覆盖了新西兰基督城的土地, 包含

了大约 2.2 万幢建筑物。区域内的影像被裁切为 8188 张图像, 其中 4736 张图像用于训练, 1036 张图像用于验证, 2416 张图像用于测试。Massachusetts 建筑物数据集由覆盖波士顿地区的 151 张航拍光学遥感影像组成, 每幅影像大小为 1500×1500 , 地面分辨率为 1m, 其包含了广泛的城市和郊区场景, 建筑物大小, 纹理, 形状各不相同。为了便于训练和评估, 此处采用边缘重叠方法将 151 张图片裁剪为 1195 张大小为 512×512 的图像, 其中 1069 张图像用于训练, 36 张图像用于验证, 90 张图像用于测试。

为了验证本文所提出网络的有效性, 采用精确度 (precision)、F1 分数 (F1-score)、召回率 (recall) 和交并比 (intersection over union, IoU) 四个评价指标来评估网络的性能。其中, 精确度与召回率分别反映预测结果的准确性与完整性, F1 分数为二者的综合指标, IoU 用于衡量预测结果与真实标注之间的重叠程度。

2.3 消融实验

为了验证所提的 MFEE-Net 网络中各模块及损失函数超参数的有效性, 选择在 WHU 建筑物数据集上进行消融实验, 并对 IoU、F1-score、precision 和 recall 等定量评价指标进行对比。

1) 模块消融实验

为了验证所设计的轻量级编码器、层间特征融合模块 IFFM 及边缘感知增强模块 EAEM 对于所提网络在性能方面的贡献, 开展如下消融实验。首先选取使用 ResNet34 作为特征提取网络的 U-Net 网络作为基线网络, 其次将 ResNet34 替换为所设计的轻量化特征提取网络, 然后在其基础上依次添加所设计的 IFFM 模块及 EAEM 模块以验证各个模块的有效性。为了保证不同网络结构对比的公平性, 所有实验均在相同的参数配置及实验环境下进行, 实验结果如表 2 所示。

可以看出, 在基线网络中, 将其主干网络 ResNet34 替换为 ResMBC Block 作为基本单元的轻量级特征提取网络后, IoU、F1-score、precision 和 recall 分别提升 0.64%、0.55%、0.97% 及 0.12%, 特别地, 网络的参数量下降了 90.59%, 说明了 ResMBC Block 通过多支路并行卷积可以有效提取多尺度特征, 增大了网络的感受野, 可以在大幅度减少网络参数的情况下提升网络的精度。其次, 在轻量级网络基础之上添加了所设计的层间特征提取模

块 IFFM, IoU、F1-score、precision 和 recall 分别提升 0.47%、0.26%、0.49% 及 0.04%, 证明了 IFFM 模块可以有效聚合编解码器不同层级的特征, 增强前景信息, 抑制背景噪声及冗余信息。在基线网络上添加边缘感知增强模块 EAEM 后, IoU、F1-score、precision 和 recall 分别提升 0.53%、0.28%、0.52% 和 0.05%, 验证了 EAEM 模块可以有效对建筑物的边缘轮廓信息进行增强监督, 从而提升网络对于建筑

物的整体提取精度。最后, 在基线网络上同时添加 IFFM 模块和 EAEM 模块构成网络的完整结构, IoU、F1-score、precision 和 recall 达到了最优值, 分别为 91.13%、95.36%、95.81% 及 94.92%, 此时的网络参数量仅为 6.39M, 证明了本文所提的融合多级特征提取与边缘增强的网络可以在更小的参数量下实现更高的建筑物提取精度。

表 2 模块消融实验结果
Table 2 Results of module ablation experiment

网络结构	IoU(%)	F1-score(%)	precision(%)	recall(%)	参数量(M)
基线网络(ResNet34)	89.57	94.31	93.95	94.67	48.54
基线网络(ResMBC)	90.21	94.86	94.92	94.79	4.57
基线网络(ResMBC)+EAEM	90.74	95.14	95.44	94.84	4.96
基线网络(ResMBC)+IFFM	90.68	95.12	95.41	94.83	6.32
基线网络(ResMBC)+EAEM+IFFM	91.13	95.36	95.81	94.92	6.39

注:加粗字体为每列最优值。

2) 损失函数消融实验

为验证所设计的联合损失函数中比例系数 α 对实验结果的影响, 在不同的比例系数配置下进行实验并对比了评价指标。实验结果如表 3 所示。

可以看出, α 的不同取值对网络的性能指标产生了显著的影响。首先, 在仅有主损失函数 \mathcal{L}_m 的情况下 (即 $\alpha = 0$), 网络的 IoU、F1-score、precision 和 recall 分别为 90.77%、95.16%、95.41% 和 94.91%。对比实验表明, 当网络仅依赖交叉熵损失和骰子损失时, 由于缺乏对于边缘的监督, 网络性能未达到最佳状态。其次, 在添加辅助损失函数 \mathcal{L}_{aux} 后, 随着 α 取值的增加, 网络精度逐渐提升, 当 $\alpha = 0.6$ 时, 网络的性能达到最优, 其 IoU、F1-score 和 precision 分别为 91.13%、95.36% 和 95.81%。说明在适当的权重分配下, 主损失函数与辅助损失函数能达到平衡, 共同提升网络对于建筑物的提取性能。

2.4 对比实验

为了评估本文提出的 MFEE-Net 网络的有效性, 与现有的主流语义分割网络在 WHU 航空影像建筑物数据集和 Massachusetts 建筑物数据集上进行了建筑物提取的对比实验。对比网络包括经典的基于 FCN 的语义分割算法: U-Net (Ronneberger 等, 2015)、DeepLabV3 (Chen 等, 2017) 和 HRNet (high-

表 3 损失函数消融实验结果

Table 3 Results of loss function ablation experiment

比例系数 α	IoU (%)	F1-score (%)	precision (%)	recall (%)
0	90.77	95.16	95.41	94.91
0.2	90.85	95.24	95.52	94.97
0.4	90.95	95.30	95.67	94.93
0.6	91.13	95.36	95.81	94.92
0.8	90.83	95.22	95.56	94.89
1.0	90.70	95.12	95.35	94.90

注:加粗字体为每列最优值。

resolution network) (Sun 等, 2019), 基于 U-Net 改进的算法: MSL-Net (multi-scale level network) (Qiu 等, 2022)、BOMSC-Net (boundary optimization and multi-scale context network) (Zhou 等, 2022)、MM-UNet (multi-module UNet) (明兴涛和杨德宏, 2024)、BEARNet (buildings edge-aware refined network) (Lin 等, 2023)、MSRF-Net (multiscale receptive field network) (Zhao 等, 2023) 和 EGAFNet (Yang 等, 2025), 以及基于 Transformer 全局信息建模的方法: Segformer (Xie 等, 2021)、UNetFormer (Wang 等, 2022) 和 MSHFormer (Zhu 等, 2025)。本文所提的 MFEE-

Net也属于基于U-Net的改进方法,因此与上述第二类方法具有较强的可比性。

1) WHU 航空影像建筑物数据集对比实验

本文所提的MFEE-Net网络与各对比网络在WHU航空影像建筑物数据集的定量实验结果如表4所示。可以看出,在WHU航空影像建筑物数据集上,本文所提网络的IoU、F1-score、precision和recall均达到了最优值,分别为91.13%、95.36%、95.81%及94.92%,与次优网络MSL-Net相比, IoU、F1-score、precision和recall分别提升了0.73%、0.36%、0.71%及0.12%,这表明,本文所提出的网络凭借良好的多尺度特征提取能力和边缘细节捕捉能力,显著提升了建筑物提取的精度,展示了该网络在遥感影像建筑物提取任务中的优越性能。

为进一步直观展示MFEE-Net在遥感影像建筑物提取任务中的实际效果,本文在WHU航空影像建筑物数据集上给出了部分可视化结果,如图5所示。在图5第一行所示场景中,重点关注区域包含大量

表4 WHU航空影像建筑物数据集定量对比实验结果

Table4 Quantitative comparison experiment results on WHU aerial imagery building dataset

对比网络	IoU (%)	F1-score (%)	precision (%)	recall (%)
U-Net	88.46	93.84	93.73	94.02
DeepLabV3	88.42	93.85	94.59	93.12
HRNet	88.71	94.02	94.41	93.62
Segformer	87.03	93.07	93.41	92.73
UNetFormer	89.66	94.57	94.82	94.34
MSL-Net	90.40	95.00	95.10	94.80
BOMSC-Net	90.15	94.80	95.14	94.50
MM-UNet	90.11	94.80	95.35	94.25
MSHFormer	89.10	94.20	94.80	93.70
Ours	91.13	95.36	95.81	94.92

注:加粗字体为每列最优值。

分布密集的小型建筑物。得益于ResMBC Block

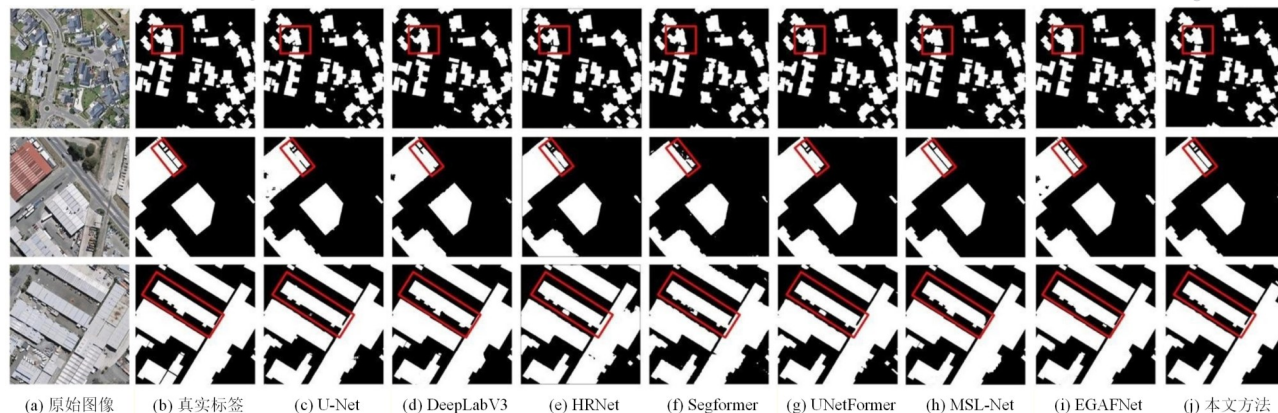


图5 WHU航空影像建筑物数据集不同网络可视化结果

Fig. 5 Visualized results of different networks on WHU aerial imagery building dataset ((a)original images; (b>true labels; (c)U-Net; (d)DeepLabV3; (e)HRNet; (f)Segformer; (g)UNetFormer; (h)MSL-Net; (i)EGAFNet; (j)ours)

的特征提取网络,所提网络能够充分挖掘不同尺度下的多层次特征,有效捕捉小型建筑物的纹理、形状等局部细节信息,从而在密集小目标场景中表现出更稳定、精细的提取效果。图5第二行场景中的重点关注区域展示了一座大型仓库及其邻近附着的小型建筑物,两者在尺度上存在显著差异。通过引入边缘感知增强模块(EAEM)以及基于边缘监督的损失函数,所提网络能够在不同尺度建筑物之间建立更加清晰的边缘约束,增强区域间的结构一致性,从而实现两类建筑物边界的准确区分。相比

之下,对比方法(如U-Net、DeepLabV3和UNetFormer)在复杂边缘感知方面仍面临一定挑战,在尺度差异较大的相邻建筑物场景中,提取结果中出现了边界不够清晰的情况。图5第三行场景展示了一个边缘轮廓较为复杂的建筑物。U-Net、HRNet以及MSL-Net在该场景中对精细边缘结构的刻画能力相对有限,难以完整保留建筑物的复杂轮廓信息。SegFormer由于采用轻量化解码器结构,在上采样过程中对细节信息的恢复能力相对受限,导致提取结果中的建筑物边缘表现较为粗略。相比之下,所提

MFEE-Net 依托其增强的边缘感知与特征细化能力,能够更完整、准确地恢复建筑物的边缘轮廓,展现出更优的视觉效果与结构一致性。

2) Massachusetts 建筑物数据集对比实验

表 5 给出了在分辨率较低的 Massachusetts 建筑物数据集上各方法的定量对比结果。可以看出,在该数据集上,尽管所提 MFEE-Net 网络的 IoU 略低于 BEARNet, recall 低于最新的方法 MSHFormer,但是在 F1-score 和 precision 指标上均取得了最优表现,分别达到 86.01% 和 87.84%,且与次优的 BEARNet 相比分别提升了 0.92% 和 2.92%,体现出更高的整体提取质量与误检抑制能力。值得注意的是, F1-score 作为综合衡量建筑物提取性能的重要指标,能够同时反映模型在召回率与提取精度之间的平衡,本文所提网络在该指标上取得最优结果,表明其在建筑物提取任务中具备更为稳健和均衡的综合性能。此外,与当前主流的基于卷积神经网络的 HRNet 以及基于 Transformer 的 SegFormer 和 UNetFormer 相比,所提方法在各项定量评价指标上均表现出明显更具竞争力的性能优势。

为直观展示所提网络在较低分辨率遥感影像上的建筑物提取能力,对 Massachusetts 数据集上的建筑物分割效果进行了可视化分析,如图 6 所示。在图 6 第一行场景的重点关注区域中,建筑物屋顶颜色与周围地面背景高度相似,且其结构形态使

表 5 Massachusetts 建筑物数据集定量对比实验结果

Table 5 Quantitative comparison experiment results of Massachusetts building dataset

对比网络	IoU (%)	F1-score (%)	precision (%)	recall (%)
U-Net	69.46	81.98	85.28	78.92
DeepLabV3	65.13	78.83	82.93	75.20
HRNet	67.70	80.74	84.34	76.67
Segformer	65.22	79.71	83.65	74.57
UNetFormer	71.83	83.60	85.33	81.95
MSL-Net	70.90	83.00	81.90	84.10
BEARNet	75.82	85.09	84.92	85.27
MM-UNet	73.42	84.72	86.39	83.21
MSHFormer	73.60	84.80	82.50	87.30
Ours	75.46	86.01	87.84	84.26

注:加粗字体为每列最优值。

得建筑物中部区域易被误判为“天井”或背景区域。在该复杂场景下,各对比方法均在不同程度上出现了局部误判,导致建筑物提取结果存在不完整的问题。相比之下,所提 MFEE-Net 网络通过引入 ResMBC Block 的特征提取网络与 IFFM 模块,显著增强了对前景与背景的判别能力,能够有效抑制背景干扰,实现对该建筑物的完整提取,体现了方法在复杂背景条件下的适应性。图 6 第二行场景中的建筑物边缘纹理复杂,屋顶颜色差异极大,U-Net、HRNet、UNetFormer 和 EGAFNet 受到不同程度的干扰,将建筑物部分区域误识别为背景。SegFormer 和 MSL-Net 虽能够较好地感知建筑物的颜色差异,但在边缘细节恢复方面仍存在不足。而本文提出的网络通过加强多尺度特征提取和边缘监督能力,能够准确提取该建筑物的完整部分及边缘轮廓。图 6 第三行场景展示了一个由小型建筑物构成的密集区域,在该场景中,由于小型建筑物形态相对规则,多数对比方法均能够较好地完成基本提取。然而,在重点关注区域内,一组屋顶颜色明显不同于周围建筑的小型建筑物群中,其余对比方法的提取结果仍与真值存在一定差距。相比之下,所提网络在该复杂局部区域中表现出更强的区分能力,能够更准确地完成该建筑物群的提取。以上可视化结果表明,所提网络在低分辨率遥感影像场景下仍能够实现对该建筑物的稳定、精细提取,具有较强的鲁棒性和泛化能力,验证了方法在不同数据条件下的有效性。

3) 建筑物边缘提取对比实验

为进一步验证本文所提 MFEE-Net 在建筑物边缘提取方面的有效性,采用拉普拉斯算子对各网络预测的建筑物分割结果进行边缘提取,并与真实建筑物边缘进行对比分析,其可视化结果如图 7 所示。所选取的测试场景中,两幅来自 WHU 航空影像建筑物数据集,一幅来自 Massachusetts 建筑物数据集,且均包含边缘结构复杂、形态多样的建筑物,具有较强的代表性和挑战性。图 7 第一行场景的重点关注区域为一幢边缘轮廓较为复杂的建筑物。U-Net、SegFormer 和 UNetFormer 在该场景中对建筑物边缘细节的刻画能力相对有限,提取结果中边缘不够清晰。HRNet 虽能够较好地识别出建筑物的整体轮廓,但在局部区域仍存在覆盖不完整的情况,导致提取结果中出现一定的空洞现象。相比之下,只有 MFEE-Net 与 MSL-Net 网络能够准确提取建筑物的边缘轮

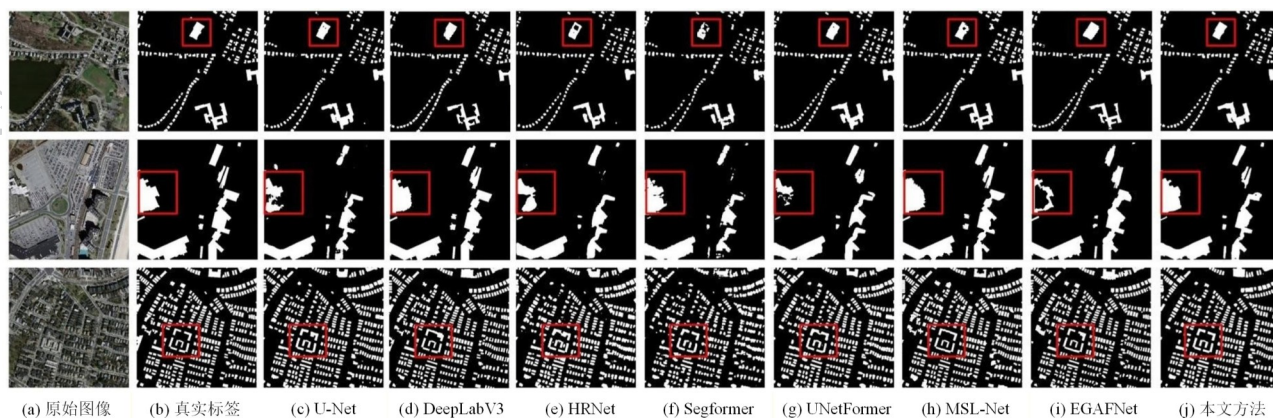


图6 Massachusetts 建筑物数据集不同网络可视化结果

Fig. 6 Visualized results of different networks on Massachusetts building dataset ((a)original images; (b>true labels; (c)U-Net; (d) DeepLabV3; (e)HRNet; (f)Segformer; (g)UNetFormer; (h)MSL-Net; (i)EGAFNet; (j)ours)

廓,且能够保持建筑物主体结构的完整性,从而更充分地保留了建筑物的细节信息。图7第二行场景的重点关注区域为一幢高层建筑,其边缘结构复杂且细节丰富。对比方法在该场景下对细粒度边缘信息的建模能力仍存在一定不足,难以全面捕捉建筑物的复杂边界特征。所提网络通过引入EAEM模块及边缘监督损失函数,在训练过程中强化了对边缘信

息的关注,从而能够更加准确地识别并提取建筑物的边缘轮廓,有效保障了分割结果的完整性与精度。图7第三行场景展示了Massachusetts数据集一个建筑物密集分布的复杂区域,其中建筑物边缘相互交错、结构关系紧密。从边缘可视化结果可以看出,所提网络在该场景中能够实现更加连续、完整的边缘提取效果。从以上分

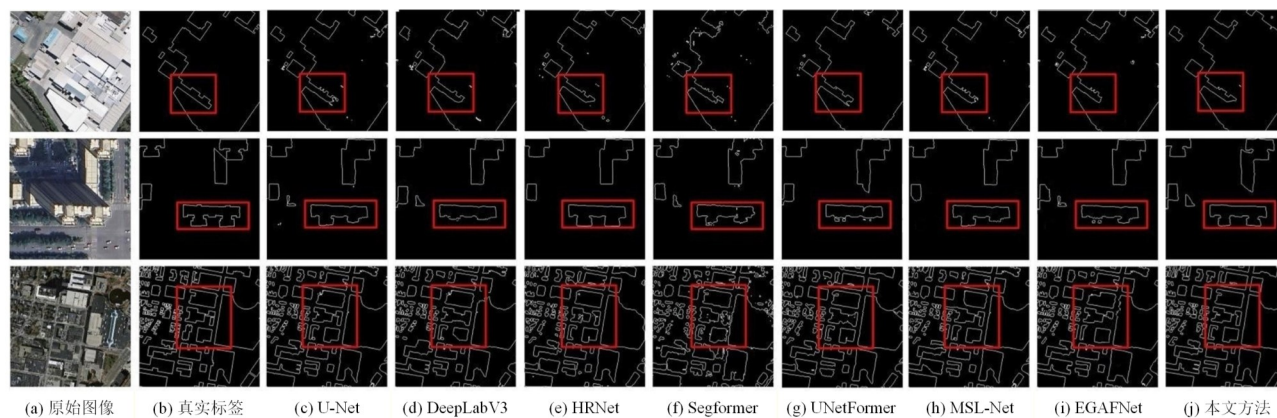


图7 建筑物边缘提取可视化结果

Fig. 7 Visualized results of building edge extraction ((a)original images; (b>true labels; (c)U-Net; (d)DeepLabV3; (e)HRNet; (f) Segformer; (g)UNetFormer; (h)MSL-Net; (i)EGAFNet; (j)ours)

析可见,本文提出的MFEE-Net网络通过引入EAEM模块及边缘损失函数,有效增强了模型对建筑物边缘轮廓信息的识别能力,从而在多种复杂场景下显著提升了建筑物边缘细节特征的提取效果。

4)模型复杂度与效率对比

为了更全面地说明本文所提MFEE-Net在模型轻量化方面的优势,我们进一步与其他方法对比了模型的参数量(单位:M)、推理速度(单位:FPS)以及

在WHU建筑物数据集上的IoU指标,如表6所示。可以看出,得益于编码器主干网络ResMBC轻量化的设计,以及EAEM模块较低的参数量,本文所提出的MFEE-Net整体的计算复杂度和效率都均优于经典的UNet、DeepLabV3和HRNet等网络,仅略低于MSL-Net,在对比方法中计算效率排名第二。在保持较高计算效率的同时,MFEE-Net的IoU指标相比MSL-Net提升了0.73%,表明该方法在仅付出极小

计算开销代价的前提下,实现了更优的分割精度表现。

表6 不同网络参数量与推理速度比较

Table. 6 Comparison of parameters and inference speed for different networks

对比网络	参数量 (M)	推理速度 (FPS)	IoU (%)
U-Net	17.26	37.12	88.46
HRNet	29.53	30.97	88.71
DeepLabV3	15.31	86.37	88.42
Segformer	7.7	73.25	87.03
UNetFormer	11.72	115.21	89.66
MSL-Net	6.0	101.44	90.40
BOMSC-Net	192.32	-	90.15
EGAFNet	44.6	25.62	89.3
MSRF-Net	12.62	-	90.16
Ours	6.39	105.32	91.13

注:加粗字体为每列最优值,“-”表示相关数据未报道。

3 结论

本文针对遥感影像中建筑物尺度变化显著、边缘轮廓易模糊等问题,提出了一种融合多级特征提取与边缘增强的建筑物提取网络 MFEE-Net。该网络基于编码器-解码器框架,通过轻量化多尺度编码结构实现对建筑物结构与纹理信息的有效建模,并借助层间特征融合模块 IFFM 加强不同尺度特征之间的协同表达;同时,在解码阶段引入边缘感知增强模块 EAEM 与边缘监督损失,有效提升了建筑物边界细节的表达能力。实验结果表明,所提方法在 WHU 和 Massachusetts 公开数据集上均取得了具有竞争力的性能,特别是在边缘细节保持方面表现出更优的效果,验证了所提网络的有效性。虽然 MFEE-Net 在处理多尺度特征与边缘轮廓方面展现出了良好的性能,但是现阶段仅考虑了建筑物边缘的先验信息,因此还有一定的提升空间,在未来的研究中,将考虑结合更多的建筑物形态的先验信息,并通过 Transformer 或者图结构建模机制引入更丰富的全局上下文信息,以进一步提升对复杂建筑结构的分割能力。

参考文献(References)

- Chen K Y, Zou Z X and Shi Z W. 2021. Building extraction from remote sensing images with sparse token transformers. *Remote Sensing*, 13 (21): #4441 [DOI: 10.3390/rs13214441]
- Chen L-C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 (4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]
- Chen X, Qiu C P, Guo W Y, Yu A Z, Tong X C and Schmitt M. 2022. Multiscale feature learning by transformer for building extraction from satellite images. *IEEE Geoscience and Remote Sensing Letters*, 19: 1-5 [DOI: 10.1109/LGRS.2022.3142279]
- Chen Z Y, Luo Y H, Wang J, Li J, Wang C and Li D L. 2023. DPENet: Dual-path extraction network based on CNN and transformer for accurate building and road extraction. *International Journal of Applied Earth Observation and Geoinformation*, 124: # 103510 [DOI: 10.1016/j.jag.2023.103510]
- Dornaika F, Moujahid A, Youssef E M and Ruichek Y. 2016. Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors. *Expert Systems with Applications*, 58: 130-142 [DOI: 10.1016/j.eswa.2016.03.024]
- Dosovitskiy A. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* [DOI: 10.48550/arXiv.2010.11929]
- Feng D J, Chen H Y, Liu S N, Liao Z Y, Shen X Y and Xie Y K, et al. 2023. Boundary-semantic collaborative guidance network with dual-stream feedback mechanism for salient object detection in optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1-17 [DOI: 10.1109/TGRS.2023.3332282]
- Fu W, Xie K and Fang L Y. 2024. Complementarity-aware local-global feature fusion network for building extraction in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-13 [DOI: 10.1109/TGRS.2024.3370714]
- Hosseinpour H, Samadzadegan F and Javan F D. 2022. CMGFNet: A deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184: 96-115 [DOI: doi.org/10.1016/j.isprsjprs.2021.12.007]
- Hu Y, Wang Z B, Huang Z and Liu Y. 2023. PolyBuilding: Polygon transformer for building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 199: 15-27 [DOI: 10.1016/j.isprsjprs.2023.03.021]
- Li K C, Wang Y L, Zhang J H, Gao P, Song G L and Liu Y, et al. 2023. Uniformer: Unifying convolution and self-attention for visual

- recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10): 12581-12600 [DOI: 10.1109/TPAMI.2023.3282631]
- Li Q Y, Mou L C, Hua Y S, Shi Y L and Zhu X X. 2021. Building footprint generation through convolutional neural networks with attraction field representation. *IEEE Transactions on Geoscience and Remote Sensing*, 0: 1-17 [DOI: 10.1109/TGRS.2021.3109844]
- Li G Y, Wu H M, Dong C H and Liu Y. 2023. Lightweight multi-scale difference network for remote sensing building extraction. *Optics and Precision Engineering*, 31(22): 3371-3382 (李国燕, 武海苗, 董春华, 刘毅. 2023. 面向遥感建筑物提取的轻型多尺度差异网络. *光学精密工程*, 31(22): 3371-3382) [DOI: 10.37188/OPE.20233122.3371]
- Li X H, Bai X C, Li Z J and Zuo Z Y. 2022. High-resolution image building extraction based on multi-level feature fusion network. *Geomatics & Information Science of Wuhan University*, 47(8): #1236 (李星华, 白星辰, 李正军, 左芝勇. 2022. 面向高分影像建筑物提取的多层次特征融合网络. *武汉大学学报(信息科学版)*, 47(8): #1236) [DOI: 10.13203/j.whugis20210506]
- Lin H J, Hao M, Luo W Q, Yu H Y and Zheng N S. 2023. BEARNet: A novel buildings edge-aware refined network for building extraction from high-resolution remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 20: 1-5 [DOI: 10.1109/LGRS.2023.3272353]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//*Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*. Boston; IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Jin S, Guan M, Bian Y C and Wang S L. 2023. Building extraction from remote sensing images based on improved U-Net. *Laser & Optoelectronics Progress*, 60(4): #0401002 (金澍, 关沫, 边玉婵, 王舒磊. 2023. 基于改进 U-Net 的遥感影像建筑物提取方法. *激光与光电子学进展*, 60(4): #0401002) [DOI: 10.3788/LOP213004]
- Ma J C, Li B, Li H X, Meng S Y, Lu R T and Mei S H. 2025. Remote sensing change detection by pyramid sequential processing with Mamba. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* [DOI: 10.1109/JSTARS.2025.3591834]
- Ming X T and Yang D H. 2024. Building extraction from remote sensing image based on multi-module. *Laser & Optoelectronics Progress*, 2024, 61(4): #0428004 (明兴涛, 杨德宏. 2024. 基于多模块的遥感影像建筑物提取方法. *激光与光电子学进展*, 61(4): #0428004) [DOI: 10.3788/LOP231148]
- Onojehuo A O, Miao Y and Blackburn G A. 2023. Deep ResU-Net convolutional neural networks segmentation for smallholder paddy rice mapping using sentinel 1 SAR and Sentinel 2 optical imagery. *Remote Sensing*, 15(6): #1517 [DOI: 10.3390/rs15061517]
- Pan J, Bai Y C, Shu Q D, Zhang Z E, Hu J R and Wang M. 2024. MSwin: Transformer-based multiscale feature fusion change detection network within cropland for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-16 [DOI: 10.1109/TGRS.2024.3374421]
- Qian K E, Ding X K, Jiang X L, Ji Y Y and Dong L. 2024. CFF-Net: Cross-hierarchy feature fusion network based on composite dual-channel encoder for surface defect segmentation. *Electronics*, 13(23): #4714 [DOI: 10.3390/electronics13234714]
- Qiu Y, Wu F, Yin J C, Liu C Y, Gong X Y and Wang A D. 2022. MSL-Net: An efficient network for building extraction from aerial imagery. *Remote Sensing*, 14(16): #3914 [DOI: 10.3390/rs14163914]
- Rong Y, Lin H and Hang R L. 2026. Boundary-guided and shape-aware semi-supervised semantic segmentation for remote sensing image [OL]. *Journal of Image and Graphics*, 1-15 (戎誉, 林辉, 杭仁龙. 边界引导与形状感知的遥感图像半监督语义分割. *中国图象图形学报*, 2026, 1-15 [DOI: 10.11834/jig.250575]
- Ronneberger O, Fischer P and Brox T. U-net: Convolutional networks for biomedical image segmentation//*Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sun K, Zhao Y, Jiang B R, Cheng T H, Xiao B and Liu D, et al. 2019. High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv: 190404514* [DOI: 10.48550/arXiv.1904.04514]
- Sun Z H, Fang H, Deng M X, Chen A J, Yue P and Di L P. 2015. Regular shape similarity index: A novel index for accurate extraction of regular objects from remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7): 3737-3748 [DOI: 10.1109/TGRS.2014.2382566]
- Turker M and Koc-San D. 2015. Building extraction from high-resolution optical spaceborne images using the integration of support vector machine (SVM) classification, Hough transformation and perceptual grouping. *International Journal of Applied Earth Observation and Geoinformation*, 34: 58-69 [DOI: 10.1016/j.jag.2014.06.016]
- Varshosaz M, Sajadian M, Pirasteh S and Moghimi A. 2024. Automated two-step seamline detection for generating large-scale orthophoto mosaics from drone images. *Remote Sensing*, 16(5): #903 [DOI: 10.3390/rs16050903]
- Wang L B, Fang S H, Meng X L and Li R. 2022. Building extraction with vision transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-11 [DOI: TGRS.2022.3186634]
- Wang L B, Li R, Zhang C, Fang S H, Duan C X and Meng X L, et al. 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190: 196-214 [DOI: 10.1016/j.isprsjprs.2022.06.008]
- Wang W L, Yu P, Li M M, Zhong X J, He Y R and Su H, et al. 2025. Tdfnet: Twice decoding v-mamba-cnn fusion features for building extraction. *Geo-spatial Information Science*, 1-20 [DOI: 10.1080/10095020.2025.2514812]

- Xia L G, Mi S L, Zhang J X, Luo J C, Shen Z F and Cheng Y B. 2023. Dual-stream feature extraction network based on CNN and transformer for building extraction. *Remote Sensing*, 15 (10) : #2689 [DOI: 10.3390/rs15102689]
- Xie E Z, Wang W J, Yu Z D, Anandkumar A, Alvarez J M and Luo P. 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34: 12077-12090 [DOI: 10.1016/j.neucom.2024.129241]
- Xie Y K, Zhan N, Zhu J, Xu B L, Chen H Y and Mao W F, et al. 2024. Landslide extraction from aerial imagery considering context association characteristics. *International Journal of Applied Earth Observation and Geoinformation*, 131: #103950 [DOI: 10.1016/j.jag.2024.103950]
- Xie Y K, Zhu J, Cao Y G, Zhang Y H, Feng D J and Zhang Y C, et al. 2020. Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features. *IEEE Access*, 8: 81904-81917 [DOI: 10.1109/ACCESS.2020.2991338]
- Yang M W, Zhao L K, Ye L F, Jia W S, Jiang H W and Yang Z. 2025. EGAFNet: an edge guidance and scale-aware adaptive fusion network for building extraction from remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1-3 [DOI: 10.1109/TGRS.2024.3524547]
- Yu S and Wang X L. 2021. Remote sensing building segmentation by CGAN with multilevel channel attention mechanism. *Journal of Image and Graphics*, 26(3) : 686-699 (余帅, 汪西莉. 2021. 含多级通道注意力机制的CGAN遥感图像建筑物分割. *中国图象图形学报*, 26(3) : 686-699) [DOI: 10.11834/jig.200059]
- Zhao M, Zhang C X, Yue P, Cai C W and Ye F H. 2025. MDA-RSM: multi-directional adaptive remote sensing mamba for building extraction. *GIScience & Remote Sensing*, 62(1) : 2568776 [DOI: 10.1080/15481603.2025.2568776]
- Zhao W F, Persello C and Stein A. Building instance segmentation and boundary regularization from high-resolution remote sensing images//*Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa: IEEE: 3916-3919 [DOI: 10.1109/IGARSS39084.2020.9324239]
- Zhou Y, Chen Z L, Wang B, Li S J, Liu H and Xu D Z, et al. 2022. BOMSC-Net: Boundary optimization and multi-scale context awareness based building extraction from high-resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-17 [DOI: 10.1109/TGRS.2022.3152575]
- Zhao Y H, Sun G Y, Zhang L, Zhang A Z, Jia X P and Han Z. 2023. MSRF-Net: multiscale receptive field network for building detection from remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1-14 [DOI: 10.1109/TGRS.2023.3282926]
- Zhu J, Zhang J B, Chen H Y, Xie Y K, Gu H C and Lian H J. 2024. A cross-view intelligent person search method based on multi-feature constraints. *International Journal of Digital Earth*, 17 (1) : #2346259 [DOI: 10.1080/17538947.2024.2346259]
- Zhu P P, Song Z C, Liu J L, Yan J Z, Luo X B and Tao Y X. 2025. MSHFormer: a multiscale hybrid transformer network with boundary enhancement for VHR remote sensing image building extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-16 [DOI: 10.1109/TGRS.2025.3545919]
- Zhu Q Q, Li Z, Zhang Y N and Guan Q F. 2020. Building extraction from high spatial resolution remote sensing images via multiscale-aware and segmentation-prior conditional random fields. *Remote Sensing*, 12(23) : #3983 [DOI: 10.3390/rs12233983]

作者简介

刘二虎,通信作者,男,副教授,主要研究方向为机器学习、计算机视觉等。E-mail:liuerhu@xauat.edu.cn

袁思杰,男,硕士研究生,主要研究方向为遥感目标检测与语义分割。E-mail:ysijie@xauat.edu.cn

李浩文,男,硕士研究生,主要研究方向为遥感变化检测。E-mail:1017027098@xauat.edu.cn

徐胜军,男,副教授,主要研究方向为图像处理、模式识别、人工智能与自动化系统等。E-mail:duplin@sina.com

胡煜,男,高级工程师,主要研究方向为遥感解译技术及应用。E-mail:huy@geovis.com.cn

杨甜甜,女,讲师,主要研究方向为图像处理。E-mail:28029@snsy.edu.cn