

集成主体边缘分离和多尺度信息提取的双分支建筑物提取网络

宋宝贵^{1,2,3}, 邵攀^{1,2,3}, 邵文^{1,2,3}, 张晓东⁴, 董婷^{1,2,3}

1. 三峡大学 水电工程智能视觉监测湖北省重点实验室, 宜昌 443002;

2. 三峡大学 计算机与信息学院, 宜昌 443002;

3. 三峡大学 先进计算中心, 宜昌 443002;

4. 武汉大学 测绘遥感信息工程国家重点实验室, 武汉 430079

摘要: 针对形状尺度多变和边界提取不够准确两个建筑物提取难题, 提出一种集成主体边缘分离和多尺度信息提取的双分支建筑物提取网络。首先, 利用解耦思想和光流技术, 设计一种主体边缘分离分支, 从而获取建筑物的主体和边缘特征, 增强对建筑物边界的表征能力。然后, 基于空洞卷积、深度可分离卷积和注意力机制, 构建一种轻量级多尺度信息提取分支, 来充分提取不同尺度的建筑物特征。最后, 借助所获取的主体和边缘特征, 给出一种主体和边缘特征辅助增强的损失函数, 优化网络训练过程。两组常用公开建筑物提取数据集上的实验结果表明, 所提出建筑物提取网络可行有效。

关键词: 遥感影像, 建筑物提取, 深度学习, U-Net, 主体边缘分离, 双分支, 多尺度, 轻量级

中图分类号: P2

引用格式: 宋宝贵, 邵攀, 邵文, 张晓东, 董婷. 2025. 集成主体边缘分离和多尺度信息提取的双分支建筑物提取网络. 遥感学报, 29(8): 2658-2670

Song B G, Shao P, Shao W, Zhang X D and Dong T. 2025. A two-branch building extraction network integrating main body and edge separation and multi-scale information extraction. National Remote Sensing Bulletin, 29(8): 2658-2670 [DOI: 10.11834/jrs.20253549]

1 引言

建筑物是人类社会生活不可或缺的载体, 及时精确地提取建筑物对城市规划、灾害救援、违章建筑检测、地理数据库更新等方面具有重要意义。随着遥感技术的快速发展, 遥感影像分辨率越来越高, 从高分辨率遥感影像上能够清晰地看到不同类型的建筑物, 使得遥感影像已成为建筑物提取的重要数据源。经过几十年的发展, 已有许多遥感建筑物提取方法和模型被提出 (Li 等, 2022; Luo 等, 2021)。根据使用基础理论不同, 这些方法大致可划分为两类, 传统建筑物提取和深度学习建筑物提取。

传统遥感建筑物提取主要是基于建筑物的光谱、纹理、轮廓、形状等信息通过人工设计特征

实现 (Li 等, 2015; 李治 等, 2023)。然而人工设计特征对影像的表征能力较弱、泛化能力不强, 导致传统建筑物提取方法在某些情况下不能取得理想的提取效果。

随着深度学习技术的快速发展, 其被引入到遥感建筑物提取领域, 已有许多基于深度学习的建筑物提取方法被提出 (Luo 等, 2021)。遥感建筑物提取本质上是一种语义分割任务, 将遥感影像中像素划分为建筑物 (前景) 与非建筑物 (背景) 两类。从深度学习技术的特点和建筑物自身的特点考虑, 遥感建筑物提取主要存在以下两种挑战: (1) 建筑物形状尺度多变、结构复杂, 导致建筑物提取有时不够完整, 且小型建筑物容易被遗漏; (2) 为提取影像的高层级特征, 深度学习需进行多

收稿日期: 2024-01-08; 预印本: 2024-08-04

基金项目: 国家自然科学基金 (编号: 41901341); 湖北省自然科学基金 (编号: 2024AFB867)

第一作者简介: 宋宝贵, 研究方向为遥感图像处理。E-mail: songbaogui@126.com

通信作者简介: 董婷, 研究方向为遥感图像处理。E-mail: dongt@ctgu.edu.cn

次下采样操作,使得影像细节信息容易丢失,容易造成建筑物边界的提取不准确。现有基于深度学习的建筑物提取方法多是以全卷积神经网络为基础,针对上述两种建筑物提取挑战进行设计。

针对第一种挑战,Zhu等(2021)提出一种多路径并行建筑物提取网络MAP-Net(Multiple Attending Path neural Network),通过对每条路径设定不同的分辨率,提取影像的多尺度信息。为有效提取多种尺度的建筑物,Sheikh等(2022)通过将空洞空间金字塔池化模块与U-Net结合提出一种端到端建筑物提取网络。Zhao等(2023)通过使用多尺度感受野特征编码器和多路径解码器,提出一种基于多尺度感受野的建筑物提取网络,来充分提取建筑物的多尺度特征。Tian等(2022)提出一种基于细化注意力金字塔的多尺度建筑物提取网络,通过在跳跃连接中使用卷积块注意力模块CBAM(Convolutional block attention module)来获取多尺度特征。陈国军等(2023)提出一种双分支架构网络,在语义信息分支上利用交叉特征融合模块聚合多尺度语义特征,在空间信息分支上通过集成空洞卷积和深度可分离卷积提取图像的多尺度空间特征,实现建筑物的有效提取。

针对建筑物边界提取困难,Sun等(2018)将主动轮廓理论和卷积神经网络CNN(Convolutional Neural Network)结合,来增强对建筑物边界的提取效果。Jung等(2022)首先通过整体嵌套边缘检测HED(Holistically-nested Edge Detection)策略在编码器阶段提取边缘特征,然后用边界增强模块优化所提取的建筑物边界。Yang等(2020)提出一种端到端双任务建筑物提取网络,一个任务用于检测建筑物,一个任务用于检测建筑物边界,通过添加边界检测任务来增强对建筑物边界的提取效果。

虽然上述建筑物提取网络各具优势,且能较为有效地完成建筑物提取任务,但这些方法主要考虑建筑物提取两个主要挑战中的一个,一定程度上影响了提取效果。为同时解决建筑物提取的多尺度问题和边界问题,Liu等(2021)将多尺度融合U-Net、区域建议网络和边缘约束多任务网络串联集成,来增强建筑物提取的性能。Zhao等(2022)将多尺度特征加权模块与边界增强模块集成,提出一种多尺度特征加权聚合和语义边界增强网络,提高建筑物提取的精度。

虽然已有少量工作同时考虑建筑物提取的多尺度问题和边界问题,但相关研究仍处在初级阶段,且现有这些方法采用单分支网络架构,在同一个网络分支中同时提取影像多尺度信息和增强建筑物边界信息,容易造成信息干扰,一定程度上影响建筑物提取性能。

基于上述分析,本文提出一种集成主体边缘分离MBES(Main Body and Edge Separation)和多尺度信息提取MIE(Multi-scale Information Extraction)的双分支(Two-branch)建筑物提取网络,记作MMT-Net。MMT-Net采用双分支架构,包括一个主体边缘分离分支MBESB(MBES branch)和一个轻量级多尺度信息提取分支LMIEB(Light MIE branch)。MBESB分支基于解耦思想,利用光流估计技术对建筑物特征进行主体和边缘分离,增强对建筑物信息的表征能力,特别是建筑物边缘信息。LMIEB分支首先设计一种注意力增强多尺度卷积模块AMCM(Attention-enhanced Multi-scale Convolutional Module),然后将其应用到网络的多个层级,来充分提取不同尺度的建筑物特征。此外,为更有效的计算建筑物提取网络的损失,给出一种主体和边缘特征辅助增强的损失函数。

通过两组公开建筑物提取数据集Inria(Maggiori等,2017)和WHU(Ji等,2019)验证本文方法MMT-Net的有效性。将其与ENet(Paszke等,2016)、SegNet(Badrinarayanan等,2017)、MMBNet(Zhang等,2022)、Refine-UNet(Qiu等,2023)和MAPNet(Zhu等,2021)5种方法进行对比。实验结果表明,所提出的MMT-Net能够有效提取建筑物的多尺度和边缘信息,在定性和定量两方面均优于5种对比方法。

2 本文方法

如图1所示,所提出的建筑物提取网络MMT-Net包括3个主要部分:主体边缘分离分支MBESB、轻量级多尺度信息提取分支LMIEB以及主体和边缘特征辅助增强的损失函数。其中MBESB和LMIEB两个分支均采用U-Net架构。

MBESB分支基于解耦思想,通过光流估计技术进行主体边缘特征分离,来提升对建筑物提取的性能。解耦思想本质上是减少系统内部的耦合程度,降低系统内部各模块(组件)之间的依赖性。当建筑物主体和边缘信息耦合度较高时,使

得建筑物信息更加复杂,不利于建筑物信息的提取,特别是建筑物的边缘区域。MBESB分支以文献(He等,2021)给出的改进U-Net(Improved U-Net, IU-Net)为基础,通过两条路径提取遥感影像中的建筑物特征(图1):第1条路径为基础特征提取路径,采用与IU-Net网络相同的操作生

成全建筑物特征 F_1 。第2条路径为主体边缘分离路径:通过将解耦思想分别应用到所生成的第3和第4层、第4和第5层特征,基于光流估计构建主体边缘分离模块MBESM(MBES module),实现对建筑物特征的主体边缘分离。

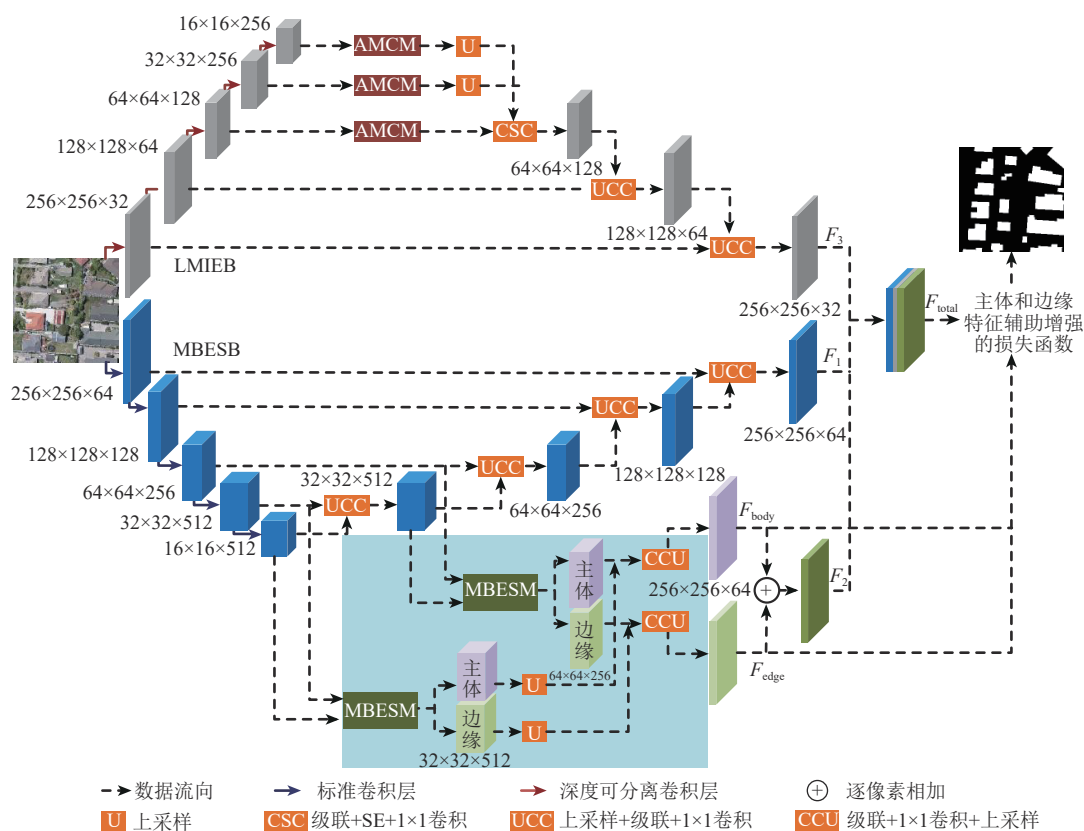


图1 MMT-Net模型结构

Fig. 1 MMT-Net model structure

LMIEB分支通过构建一种多层次多尺度网络来提取不同尺度的建筑物特征 F_3 :首先,为提取不同尺度的特征,设计一种注意力增强多尺度卷积模块AMCM;然后将AMCM应用到网络编码生成的后三层级特征,来提取不同层级特征的多尺度特征。

所给出的主体和边缘特征辅助增强损失函数同时考虑3种类型特征:全建筑物特征 F_{total} 、建筑物主体特征 F_{body} 和建筑物边缘特征 F_{edge} 。

2.1 主体边缘分离分支MBESB

由于在编码和解码过程中,容易造成影像细节信息丢失,导致建筑物边缘的提取效果有时不够理想。针对此问题,本文基于解耦思想、利用

光流估计技术(Li等,2020),设计主体边缘分离分支MBESB对建筑物特征进行主体和边缘分离,借助分离后的边缘特征增强对建筑物边缘的提取效果。MBESB分支以IU-Net(He等,2021)网络为基础,在编码阶段共生成5个层级特征:首先通过两个卷积层对输入影像进行初步处理,得到尺寸为256x256x64的第1层级特征图;然后经过四次下采样,依次得到尺寸为128x128x128、64x64x256、32x32x512和16x16x512的第2至第5层级特征。更多编码细节信息可参见文献(He等,2021)。

通过综合考虑效率和性能,将解耦思想分别应用到所生成的第3层和第4层(此处第4层特征为上采样后的第5层特征和原第4层特征融合得到

的新特征)、第4层和第5层特征,通过光流估计技术构建主体边缘分离模块MBESM,实现建筑物信息解耦。MBESM模块的输入包含两个相邻层级(比如第4层和第5层)的特征:较高层级特征的空间分辨率较低,包含更多的低频信息特征,对应包含更多的对象主体部分,故可以将较高层级特征作为种子点,指导生成主体特征。为方便叙述,较高层级特征记作 F_H ,较低层级特征记作 F_L 。

如图2所示,主体边缘分离模块MBESM首先利用光流估计技术,借助较高层级特征 F_H ,生成较低层级特征 F_L 的主体特征 F_{body}^L ;然后通过对 F_L 和 F_{body}^L 作差,生成 F_L 的边缘特征 F_{edge}^L ,即 $F_{edge}^L = F_L - F_{body}^L$ 。其中 H 和 L 的取值包括两种情况: $H=4, L=3$ 或 $H=5, L=4$ 。

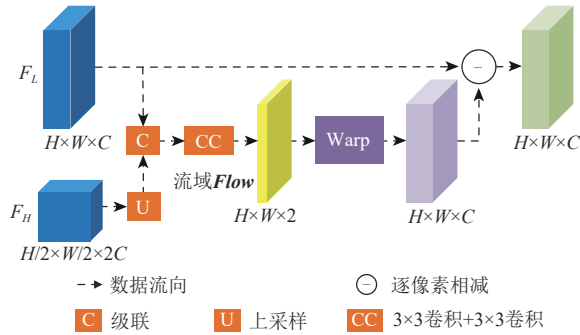


图2 主体边缘分离模块

Fig. 2 Main body and edge separation module

获取主体特征 F_{body}^L ($L=3$ 或 4)的具体过程如下:(1)将较高层级特征 F_H 进行上采样,并将上采样后的较高层级特征与较低层级特征 F_L 进行级联。(2)借助光流估计技术,基于级联后特征生成流域 $Flow$ 。 $Flow$ 本质上为通道数为2的三维矩阵,空间维尺寸与特征 F_L 相同,两个通道维度的数值表示像素在空间维的流向矢量。(3)借助流域 $Flow$ 对较低层级特征 F_L 进行变形(Warp),生成 F_L 的主体特征 F_{body}^L 。借鉴文献(Dosovitskiy等, 2015),采用如下方式生成 $Flow$:

$$Flow = Conv_{3 \times 3} \left(Conv_{3 \times 3} \left(Cat \left(F_L, Up(F_H) \right) \right) \right) \quad (1)$$

式中, Cat 表示级联, $Conv_{3 \times 3}$ 表示卷积核为 3×3 的卷积, Up 表示双线性插值上采样。与文献(Dosovitskiy等, 2015)中方法相比,本文生成 $Flow$ 的步骤中在级联后增加一个 3×3 卷积层,以便级联特征充分融合后,再生成流域 $Flow$ 。

得到流域 $Flow$ 后,首先借助其给出的像素流

向信息对特征流向进行引导,通过 $p+Flow(p)$ 将标准空间格网中每个位置的点 p 映射为新点 p^* ;然后采用与Li等(2020)相同方式对较低层级特征 F_L 进行变形(Warp),生成其主体特征 F_{body}^L :

$$F_{body}^L(p^*) = \sum_{p \in N(p)} w_p F_L(p) \quad (2)$$

式中, $N(p)$ 表示像素 p 的邻域,权重 w_p 通过流域 $Flow$ 生成。

MBESB分支生成5个层级特征,两个层级的主体特征 F_{body}^L 和边缘特征 F_{edge}^L ($L=3$ 或 4)后,通过2条解码路径对特征进行重构(图1):第1条路径为基础解码路径,采用与IU-Net相同的操作,基于5个层级特征生成全建筑物特征 F_1 。第2条路径对主体和边缘特征进行重构:首先通过式(3)对两个层级的主体特征进行融合,生成融合主体特征 F_{body} :

$$F_{body} = Up \left(Conv_{1 \times 1} \left(Cat \left(Up(F_{body}^4), F_{body}^3 \right) \right) \right) \quad (3)$$

式中, Up 表示双线性插值上采样, $Conv_{1 \times 1}$ 表示卷积核为 1×1 的卷积, Cat 表示级联。采用相同的步骤,基于 F_{edge}^3 和 F_{edge}^4 生成融合边缘特征 F_{edge} 。

如图1所示,得到融合主体特征 F_{body} 和边缘特征 F_{edge} 后,通过两种方式对其进行利用,从而增强建筑物提取效果:(1)通过式(4)将其融合,生成富含边缘信息的全建筑物特征 F_2 ,为生成最终的全建筑物特征 F_{total} 做准备;(2)基于 F_{body} 和 F_{edge} 生成建筑物主体预测概率图和建筑物边缘预测概率图,用于计算所提出主体和边缘特征辅助增强损失函数。

$$F_2 = F_{body} + F_{edge} \quad (4)$$

2.2 轻量级多尺度信息提取分支LMIEB

由于建筑物形状尺度多变,只使用同一尺度特征,不能很好地从影像中提取建筑物信息。为此,本文设计一种注意力增强多尺度卷积模块AMCM,然后将其分别应用到U型网络的后3个层级,提出一种轻量级多尺度信息提取分支LMIEB。LMIEB主要通过以下两方面实现多尺度信息的提取:(1)在AMCM中使用3种空洞率的空洞卷积获取3种尺度的信息;(2)将AMCM应用到LMIEB的多个编码层级,获取多个层级特征的多尺度信息。LMIEB主要通过3个方面实现轻量化设计:1)在编码阶段,采用计算量较小的深度可分离卷积,同时通过降低特征通道数减少卷积核个数。2)在AMCM模块

中,采用计算量较小的空洞卷积。3) 只将 AMCM 模块应用到网络的后 3 个层级,而不是所有 5 个层级。LMIEB 分支的具体细节如下:

如图 3 所示, AMCM 模块首先通过 1×1 卷积(卷积核个数为 64)进行通道降维,生成通道数为 64 的特征 F ; 其次通过 5 个分支分别对 F 进行处理: 前 2 个分支用来生成通道注意力图 W , 后 3 个分支用来生成 3 种尺度的特征 D_1 、 D_2 、 D_3 ; 然后用通道注意力图 W 来自适应增强 D_1 、 D_2 、 D_3 ; 最后对增强后特征进行融合, 得到融合特征 $\hat{F}$ 。通道注意力图 W 的生成过程如下:

$$W = \delta(GA(F) + GM(F)) \quad (5)$$

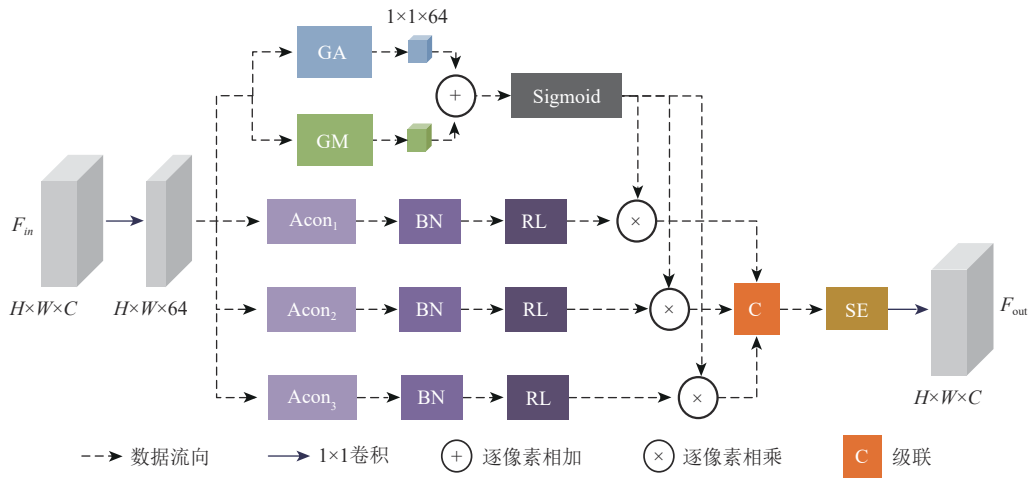


图3 注意力增强多尺度卷积融合模块

Fig. 3 Attention enhanced multi-scale convolutional fusion module

不同于经典空间金字塔池化和空洞空间金字塔池化等多尺度模块, 所提出的 AMCM 模块通过引入通道注意力机制, 来增强对多尺度特征的提取融合效果。同时, LMIEB 将 AMCM 应用到第 3 至 5 层特征, 以期更有效地提取多种尺度的建筑物信息, 从而提高不同尺度建筑物的完整性。

LMIEB 分支生成 3 个层级的多尺度特征 S_5 、 S_4 和 S_3 , 通过式 (8) 对其融合得到融合多尺度特征 S (图 1)。

$$S = \text{Fuse}(\text{Cat}(\text{Up}_1(S_5), \text{Up}_2(S_4), S_3)) \quad (8)$$

式中, Up_1 表示双线性插值四倍上采样, Up_2 表示双线性插值两倍上采样, Cat 表示级联, Fuse 为融合操作, 具体细节如式 (9) 所示:

$$\text{Fuse}(x) = \text{RL}(\text{BN}(\text{Conv}_{1 \times 1}(\text{SE}(x)))) \quad (9)$$

式中, SE 表示经典的挤压激励通道注意力,

式 (6) 中, GA 表示全局平均池化, GM 表示全局最大池化, δ 表示 Sigmoid 激活函数。第 i 种尺度的特征图 D_i 通过式 (6) 得到, $i=1, 2, 3$:

$$D_i = \text{RL}(\text{BN}(\text{ACon}_i(F))) \quad (6)$$

式中, ACon_i 表示卷积核为 3×3 空洞率为 i 的空洞卷积(卷积核个数为 64), $i=1, 2, 3$; BN 表示批量标准化, RL 表示 ReLU 激活函数。融合特征 $\hat{F}$ 的获取过程如下:

$$\hat{F} = \text{Conv}_{1 \times 1}(\text{SE}(\text{Cat}(W \times D_1, W \times D_2, W \times D_3))) \quad (7)$$

式中, SE 表示经典的挤压激励通道注意力, $\text{Conv}_{1 \times 1}$ 表示 1×1 卷积, Cat 表示级联。

$\text{Conv}_{1 \times 1}$ 表示 1×1 卷积, 卷积核个数为 C (第 3—5 编码层级中 C 的取值分别为 128、256 和 256); BN 表示批量标准化, RL 表示 ReLU 激活函数。

得到融合多尺度特征 S 后, 采用与经典 U-Net 相同的方式, 将其依次与 LMIEB 的第 2 层和第 1 层级特征进行融合, 得到尺度信息更加丰富的全建筑物特征 F_3 。然后将 F_3 与 MBESB 分支生成的全建筑物特征 F_1 和 F_2 , 采用与式 (9) 相同的方式进行融合, 得到最终的全建筑物特征 F_{total} ; 最后将 F_{total} 经过卷积层和 Sigmoid 激活函数生成全建筑物预测概率图。

2.3 主体和边缘特征辅助增强损失函数

受 Li 等 (2020) 启发, 借助 2.2 节生成的主体特征 F_{body} 和边缘特征 F_{edge} , 针对遥感建筑物提取, 提出一种主体和边缘特征辅助增强的损失函数,

来优化建筑物提取网络的训练过程。所提出的损失函数 $Loss$ 包括 3 部分: 基于全建筑物特征 F_{total} 的全损失 L_{total} 、基于建筑物主体特征 F_{body} 的主体损失 L_{body} 和基于建筑物边缘特征 F_{edge} 的边缘损失 L_{edge} , 具体如式 (10) 所示:

$$Loss = \lambda_1 L_{total} + \lambda_2 L_{body} + \lambda_3 L_{edge} \quad (10)$$

式中, λ_1 、 λ_2 和 λ_3 是 3 个权重系数, 且满足 $\lambda_1 + \lambda_2 + \lambda_3 = 1$ 。

为增强损失函数的鲁棒性, 3 种损失 L_{total} 、 L_{body} 和 L_{edge} 均通过集成二元交叉熵 BCE (Binary Cross Entropy) 和 Dice 两种经典损失函数计算:

$$L_i = \alpha_1 L_{BCE}(Pr_i, Lab_i) + \alpha_2 L_{Dice}(Pr_i, Lab_i) \quad (11)$$

式中, 下标 $i=total$ 、 $body$ 或 $edge$; Pr 表示建筑物提取的预测概率图, Lab 表示真值标签图, 比如 Pr_{body} 表示建筑物主体部分的预测概率图, Lab_{body} 表示建筑物主体部分的真值标签; L_{BCE} 和 L_{Dice} 分别表示二元交叉熵和 Dice 损失函数; α_1 和 α_2 表示两个权重系数, 且满足 $\alpha_1 + \alpha_2 = 1$ 。

3 实验结果与分析

3.1 实验数据与实验设计

为验证所提出建筑物提取网络 MMT-Net 的性能, 采用两组常用公开建筑物提取数据集进行实验, 即 Inria 数据 (Maggiori 等, 2017) 和 WHU 数据 (Ji 等, 2019)。

Inria 数据由航拍影像构成, 包括 180 张具有真值标签的训练集和 36 张不具有真值标签的测试集, 每张影像大小为 5000×5000 像素, 空间分辨率为 0.3 m 。数据涵盖不同的城市区域, 包括人口密集的中心城区和人口相对稀少的高山城镇, 覆盖区域总面积为 810 km^2 。本文使用具有真值标签的训练集对模型进行训练和测试, 共包括 5 个不同的地区, 即 Tyrol、Austin、Chicago、Kitsap 和 Vienna, 每个地区包含 36 张具有真值标签的影像。

为了充分验证 MMT-Net 网络的有效性和鲁棒性, 针对 Inria 数据的实验包括两部分: 第 1 部分对 Inria 5 个地区所有数据进行实验, 其中 155 张影像作为训练集, 25 张影像作为测试集; 第 2 部分选取 3 个典型地区 (即 Tyrol、Austin 和 Chicago) 分别进行实验, 每个地区的训练集包括 31 张影像, 测试集包括 5 张影像。

WHU 数据包括两种类型的数据集, 即卫星和航拍影像数据集。本文采用后者进行实验, 共包括 8188 张大小为 512×512 像素的航拍影像, 其中训练集 (包括 130500 个建筑物)、验证集 (包括 14500 个建筑物) 和测试集 (包括 42000 个建筑物) 分别包括 4736、1036 和 2416 张影像。WHU 数据影像的获取位置为新西兰某地区, 覆盖区域总面积为 450 km^2 。影像原始空间分辨率为 0.075 m , 数据提供者对原始影像进行重采样, 重采样后影像空间分辨率为 0.3 m 。

考虑到 GPU 限制, 将所有影像无重叠裁剪成 256×256 像素影像块。对于 Inria 数据中影像, 首先通过补 0 操作将其扩展至 5120×5120 像素大小, 然后进行裁剪。在进行训练前, 将不包含建筑物的影像块删除。

为保持模型训练在不同数据上的一致性, 对所提出主体和边缘特征辅助增强损失函数的权重系数 λ_1 、 λ_2 、 λ_3 、 α_1 和 α_2 , 本文通过 WHU 数据集的验证集训练得到, 并对 Inria 和 WHU 数据采用相同的值进行实验, 即 $\lambda_1=0.5$ 、 $\lambda_2=0.3$ 、 $\lambda_3=0.2$ 、 $\alpha_1=0.8$ 和 $\alpha_2=0.2$ 。对超参数 λ_1 、 λ_2 、 λ_3 、 α_1 和 α_2 的消融实验请参见 3.3.2 节。

采用建筑物提取文献 (Cai 和 Chen, 2021) 中常用的 4 种精度指标进行定量分析, 即精确率 Precision、召回率 Recall、F1 值和交并比 IoU。借鉴文献 (Cai 和 Chen, 2021) 中对比实验的组织方式, 设计如下对比实验: (1) 与经典语义分割网络 ENet (Paszke 等, 2016) 和 SegNet (Badrinarayanan 等, 2017) 进行对比; (2) 与最近提出的建筑物提取网络 MMB-Net (Zhang 等, 2022)、Refine-UNet (Qiu 等, 2023) 和 MAP-Net (Zhu 等, 2021) 进行对比。

论文中所有实验基于 Pytorch 框架实现, 使用的显卡型号为 NVIDIA GeForce RTX 3060, 容量为 12 G 。训练过程中, 迭代次数设置为 200, 初始学习率为 0.0001 , 训练 Batch Size 为 4, 验证 Batch Size 设置为 2, 通过 Adam 优化器迭代优化损失函数。

3.2 实验结果分析

3.2.1 Inria 数据实验结果分析

图 4 给出 Inria 数据的 5 组典型建筑物提取结果图: 图 4 (a) 为原始影像, 图 4 (b) 为真值标签,

图4 (c)–(h) 分别为 ENet、SegNet、MMB-Net、Refine-UNet、MAP-Net 和本文方法 MMT-Net 的建筑物提取结果图。其中红色区域表示虚检错误，绿色区域表示漏检错误。

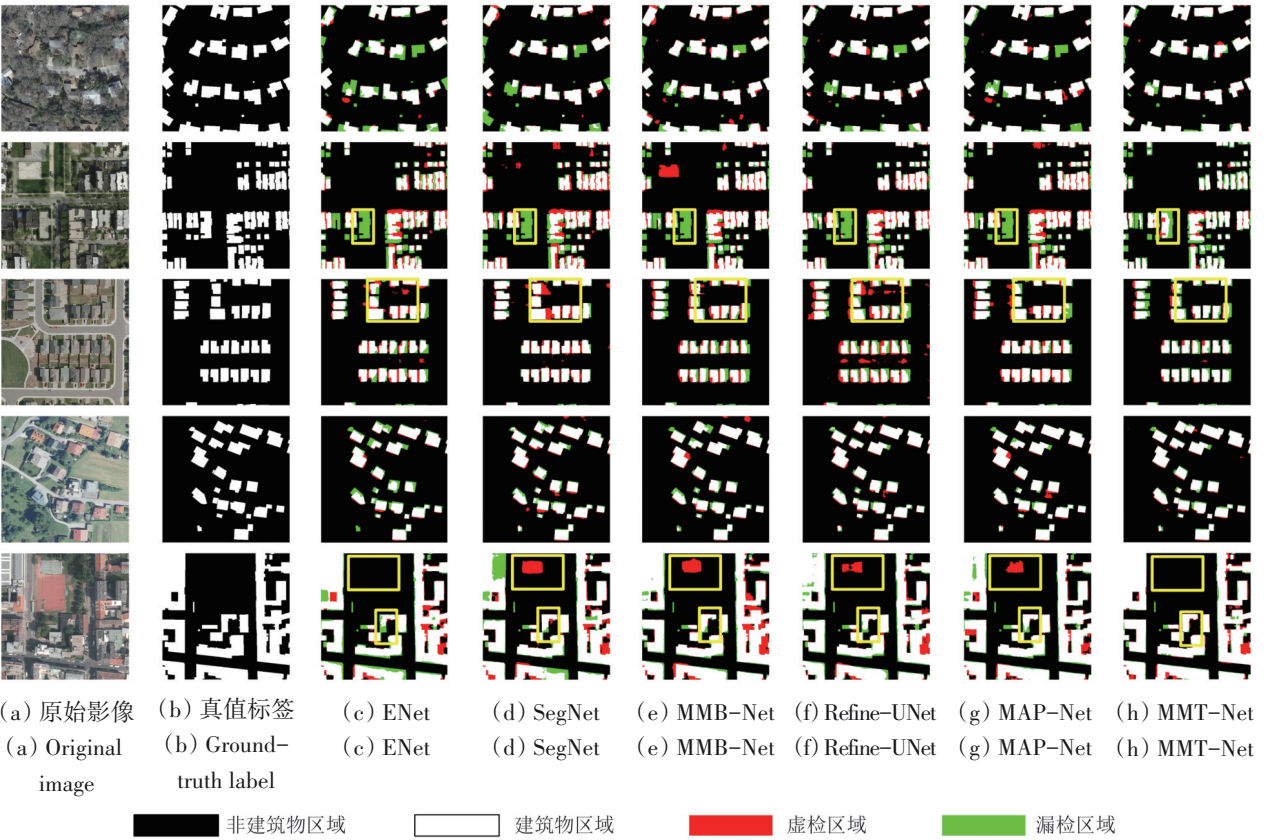


图4 Inria数据集上不同网络的可视化结果

Fig. 4 Comparison of different network visualization results in the Inria dataset

从图4可知，对于Inria数据，本文方法MMT-Net得到与真值标签最接近的建筑物提取图：MMT-Net对影像中不同密集程度、不同结构以及不同尺度的建筑物，均能比较完整地检测出，且边界较为准确；而5种对比网络表现则不够理想：图中通过黄色矩形框标出一些典型区域，比如第5行中：较大的矩形框内SegNet、MMB-Net、Refine-UNet和MAPNet有明显的虚检错误，较小的矩形框内ENet有较多的漏检错误，而本文方法MMT-Net在两个矩形框内都取得比较理想的检测结果。综上所述，MMT-Net具有较好的建筑物提取效果。

为更加客观地评价不同建筑物提取方法的性能，表1给出Inria数据集上建筑物提取结果的定量评价指标。从表1可知，在Inria数据集上，本文方法MMT-Net关于Precision、Recall、F1和IoU这4个指标均取得最优的结果。比如，MMT-Net的IoU值为0.8008，比其他方法至少提高2.67%。

表1 Inria数据集总数据对比结果

Table 1 Comparison results of total Inria dataset

方法	Precision	Recall	F1	IoU
ENet(2016)	0.8537	0.8434	0.8485	0.7369
SegNet(2017)	0.8559	0.8580	0.8570	0.7497
MMB-Net(2022)	0.8630	<u>0.8668</u>	0.8649	0.7620
Refine-UNet(2023)	0.8740	0.8636	0.8688	0.7680
MAP-Net(2021)	<u>0.8877</u>	0.8581	<u>0.8727</u>	<u>0.7741</u>
MMT-Net	0.8901	0.8889	0.8894	0.8008

注：加粗数字代表最优值，加下划线数字代表次优值。

表2给出不同方法在Tyrol、Austin和Chicago 3个地区建筑物提取结果的定量指标。从表2可知，对于3个地区，本文提出的MMT-Net在4个评价指标上均取得最优的结果。比如，MMT-Net在Tyrol、Austin和Chicago地区的IoU值分别为0.8299、0.8117和0.7444，比其他方法分别至少提高2.4%、2.29%和1.58%。

表2 Inria数据集分区对比结果

Table 2 Comparison results of Inria dataset by region

方法	Tyrol				Austin				Chicago			
	Precision	Recall	F1	IoU	Precision	Recall	F1	IoU	Precision	Recall	F1	IoU
ENet(2016)	0.8833	0.8653	0.8742	0.7765	0.8659	0.8605	0.8632	0.7593	0.8068	0.8367	0.8215	0.6970
SegNet(2017)	0.8840	0.8755	0.8797	0.7852	0.8751	0.8750	0.8750	0.7778	0.8201	0.8379	0.8289	0.7078
MMB-Net(2022)	0.8984	0.8745	0.8862	0.7957	0.8801	0.8705	0.8753	0.7782	0.8370	<u>0.8491</u>	<u>0.8430</u>	<u>0.7286</u>
Refine-UNet(2023)	0.9016	0.8789	0.8901	0.8020	0.8751	<u>0.8790</u>	0.8771	0.7811	0.8325	0.8406	0.8314	0.7215
MAP-Net(2021)	<u>0.9024</u>	<u>0.8828</u>	<u>0.8925</u>	<u>0.8059</u>	<u>0.8875</u>	0.8764	<u>0.8820</u>	<u>0.7888</u>	<u>0.8417</u>	0.8244	0.8330	0.7138
MMT-Net	0.9035	0.9106	0.9070	0.8299	0.8955	0.8966	0.8961	0.8117	0.8568	0.8516	0.8542	0.7444

注：加粗数字代表最优值，加下划线数字代表次优值。

3.2.2 WHU 数据实验结果分析

图5给出WHU数据的5组典型建筑物提取结果图：图5(a)为原始影像、图5(b)为真值标签，图5(c)—(h)分别为ENet、SegNet、MMB-

Net、Refine-UNet、MAP-Net以及本文方法MMT-Net的建筑物提取结果图。其中，红色部分表示虚检错误，绿色部分表示漏检错误。

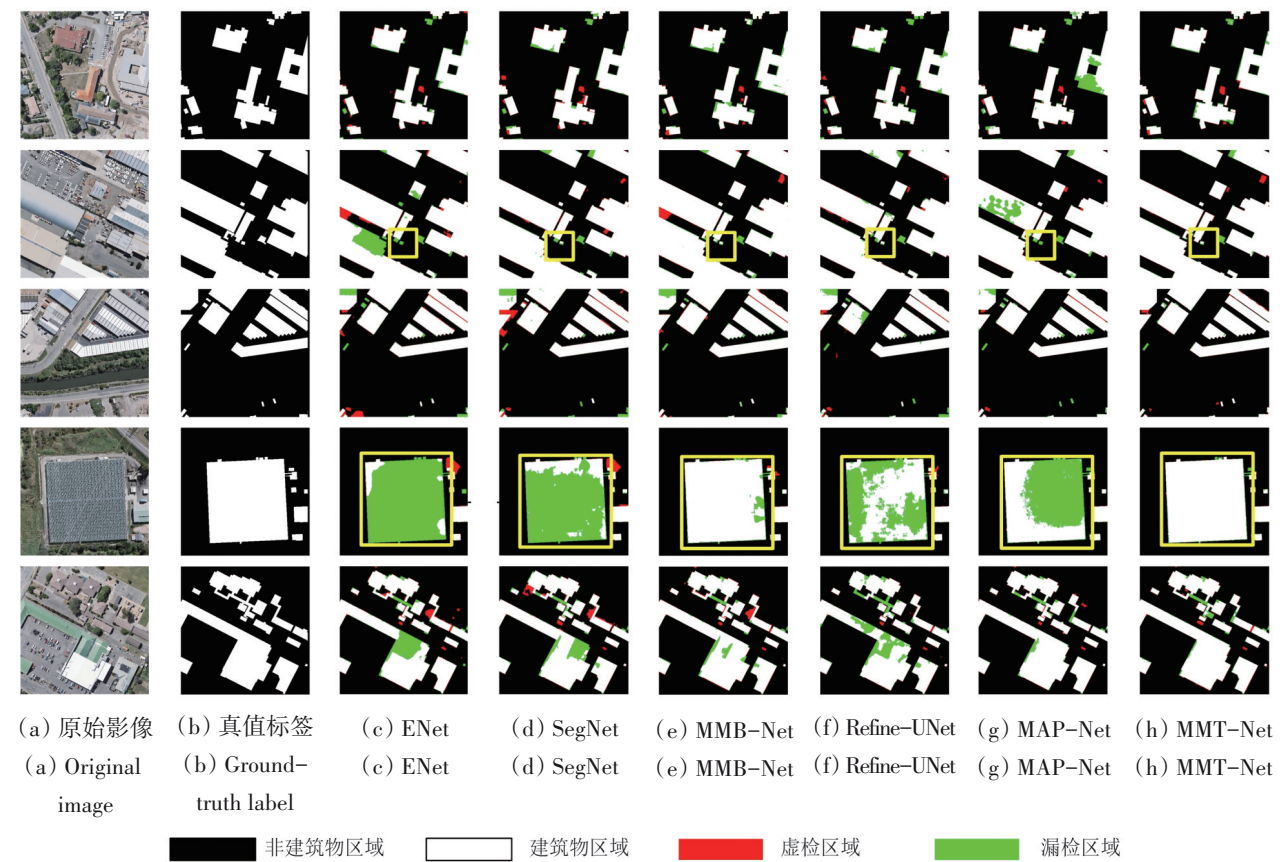


图5 WHU数据集上不同网络的可视化结果

Fig. 5 Comparison of different network visualization results in the WHU dataset

从图5可知，对WHU数据集，本文方法MMT-Net亦得到最优的建筑物提取结果：比如对第2行中矩形框内的小型建筑物，ENet和MAP-Net全部遗漏，SegNet、MMB-Net和Refine-UNet也有明显的漏检，而所提出的MMT-Net检测效果更加

理想；对于第4行矩形框内的大型建筑物，本文方法MMT-Net的检测效果明显优于其他对比方法。

表3给出WHU数据集上建筑物提取结果的定量指标。从表3可知，对WHU数据，本文方法MMT-Net也取得最优的定量指标。与其他5种方

法相比，本文方法 MMT-Net 在 4 组指标上均取得更优的结果。比如，MMT-Net 的 IoU 值为 0.9170，比其他方法至少提高 1.28%。

表 3 WHU 数据集对比结果
Table 3 Comparison results of WHU dataset

方法	Precision	Recall	F1	IoU
ENet(2016)	0.9399	0.9455	0.9427	0.8916
SegNet(2017)	0.9386	0.9410	0.9398	0.8864
MMB-Net(2022)	0.9452	0.9450	0.9451	0.8959
Refine-UNet(2023)	0.9487	0.9439	0.9463	0.8981
MAP-Net(2021)	<u>0.9493</u>	<u>0.9501</u>	<u>0.9497</u>	<u>0.9042</u>
MMT-Net	0.9581	0.9553	0.9567	0.9170

注：加粗数字代表最优值，加下划线数字代表次优值。

表 4 网络结构消融实验结果
Table 4 Ablation experiment's results of MMT-Net's structure

Baseline	LMIEB	MBESB	主体和边缘特征辅助增强损失函数	Inria				WHU			
				Precision	Recall	F1	IoU	Precision	Recall	F1	IoU
√				0.8759	0.8745	0.8752	0.7781	0.9487	0.9444	0.9466	0.8985
√	√			0.8791	0.8868	0.8829	0.7904	0.9501	0.9547	0.9524	0.9091
√		√	√	0.8864	0.8784	0.8824	0.7895	0.9563	0.9471	0.9517	0.9078
√	√	√		0.8710	0.8849	0.8779	0.7923	0.9544	0.9520	0.9532	0.9106
√	√	√	√	0.8901	0.8889	0.8894	0.8008	0.9581	0.9553	0.9567	0.9170

注：加粗数字代表最优值。

对于 MBESB 分支，通过对比表 4 中第 1 行和第 3 行的结果可知，MBESB 使得 4 个评价指标在两组数据集上均有所提升，验证了该分支的有效性。

对于主体和边缘特征辅助增强损失函数，通过对比表 4 中第 4 行和第 5 行的结果得知，本文给出的主体和边缘特征辅助增强损失函数能够有效提升模型性能，比如，所提出损失函数在 Inria 数据集上使得 F1 值和 IoU 值分别提高 1.15% 和 0.85%。

通过以上消融实验可以发现，本文设计的 MBESB、LMIEB 以及主体和边缘特征辅助增强损失函数均能提高对建筑物的提取性能。

为更加直观的展示所提出网络 MMT-Net 对于建筑物边界的提取效果，图 6 给出从 Inria 和 WHU 两个数据集中选取的 4 组典型影像的建筑物提取结果：图 6 (a)–(d) 分别为原始影像、真值标签、不包含 MBESB 分支的提取结果图和包含 MBESB 分支的提取结果图。红色部分表示虚检错误，绿色部分表示漏检错误。对比图 6 的 (c) 和 (d) 列可知：图 6 (d) 列中建筑物边界附近的彩色区域面积明显少于图 6 (c) 列中的彩色区域面积，表明 MBESB 分支能够有效增强对建筑物边界的提取效果。

3.3 消融实验

3.3.1 网络结构消融实验

本节对所提出的主体边缘分离分支 MBESB、轻量级多尺度信息提取分支 LMIEB，以及主体和边缘特征辅助增强的损失函数进行消融实验。表 4 给出消融实验结果。Baseline 表示 He 等 (2021) 给出的 IU-Net 网络。

对于 LMIEB 分支，通过对比表 4 中第 1 行和第 2 行的结果可知，LMIEB 分支使得所提出网络在 Inria 和 WHU 两组数据集上 F1 值分别提高 0.77% 和 0.58%，IoU 值分别提高 1.23% 和 1.06%，说明 LMIEB 分支能够有效提升模型的性能。

3.3.2 超参数消融实验

为测试超参数 λ_1 、 λ_2 、 λ_3 、 α_1 和 α_2 对所提出网络 MMT-Net 的影响，在综合考虑各超参数性质和计算成本的前提下，组织如下消融实验：首先通过初步实验，初步确定超参数的值；然后 (1) 固定超参数 α_1 和 α_2 的值，测试 λ_1 、 λ_2 和 λ_3 对 MMT-Net 的影响；(2) 固定超参数 λ_1 、 λ_2 和 λ_3 的值，测试 α_1 和 α_2 对 MMT-Net 的影响。

测试超参数 λ_1 、 λ_2 和 λ_3 时，固定 α_1 和 α_2 的值分别为 0.8 和 0.2。 λ_1 、 λ_2 和 λ_3 分别表示全建筑物特征损失、建筑物主体特征损失和建筑物边缘特征损失的权重；考虑到全建筑物特征同时包含建筑物主体特征和边缘特征，故令 λ_1 的权重占总体的一半左右比较合理，因此令 λ_1 在集合 {0.4、0.5、0.6} 中取值。建筑物主体特征的信息通常多于边缘特征，故可令 $\lambda_2 \geq \lambda_3$ ；同时考虑到 $\lambda_1 \in \{0.4、0.5、0.6\}$ 和 $\lambda_1 + \lambda_2 + \lambda_3 = 1$ ；所以本文在以下六组集合上测试超参数 λ_1 、 λ_2 和 λ_3 ({ λ_1 、 λ_2 、 λ_3 })：{0.4、0.4、0.2}、{0.4、0.3、0.3}、{0.5、0.4、0.1}、{0.5、0.3、0.2}、{0.6、0.3、0.1}、{0.6、0.2、0.2}。表 5 给出超参数 λ_1 、 λ_2 和 λ_3 的消融实验结果。



图6 4组典型影像的建筑物提取结果

Fig. 6 Building extraction results from four typical images

表5 超参数 λ_1 、 λ_2 和 λ_3 的消融实验结果

Table 5 Ablation experiment's results of hyper-parameter $\lambda_1, \lambda_2, \lambda_3$

λ_1	λ_2	λ_3	Inria				WHU			
			Precision	Recall	F1	IoU	Precision	Recall	F1	IoU
0.4	0.4	0.2	0.8825	0.8934	0.8879	0.7984	0.9508	0.9582	0.9545	0.9130
0.4	0.3	0.3	0.8847	0.8900	0.8873	0.7975	0.9496	0.9581	0.9538	0.9118
0.5	0.4	0.1	0.8898	0.8825	0.8861	0.7955	0.9528	0.9579	0.9553	0.9145
0.5	0.3	0.2	0.8901	0.8889	0.8894	0.8008	0.9581	0.9553	0.9567	0.9170
0.6	0.3	0.1	0.8847	0.8853	0.8850	0.7937	0.9516	0.9569	0.9543	0.9126
0.6	0.2	0.2	0.8881	0.8853	0.8867	0.7965	0.9520	0.9576	0.9548	0.9135

注：加粗数字代表最优值。

从表5可知，(1) 超参数 λ_1 、 λ_2 和 λ_3 的取值对所提出网络MMT-Net的影响较小，表明MMT-Net关于超参数 λ_1 、 λ_2 和 λ_3 具有一定的鲁棒性。(2) 当 $\lambda_1=0.5$ ， $\lambda_2=0.3$ ， $\lambda_3=0.2$ 时，MMT-Net在两组数据

集上均得到最优的建筑物提取结果。

测试超参数 α_1 和 α_2 时, 固定 λ_1 、 λ_2 和 λ_3 的值分别为 0.5、0.3 和 0.2。经过初步实验发现, 当 $\alpha_1 \geq \alpha_2$ 时, MMT-Net 能够取得更优的结果; 同时考虑到 $\alpha_1 + \alpha_2 = 1$, 故在以下 5 组集合上测试 α_1 和 α_2 ($\{\alpha_1, \alpha_2\}$): $\{0.5, 0.5\}$ 、 $\{0.6, 0.4\}$ 、 $\{0.7, 0.3\}$ 、

$\{0.8, 0.2\}$ 、 $\{0.9, 0.1\}$ 。表 6 给出超参数 α_1 和 α_2 的消融实验结果。从表 6 可知, (1) 超参数 α_1 和 α_2 的取值对所提出网络 MMT-Net 的影响较小, 表明 MMT-Net 关于超参数 α_1 和 α_2 具有一定的鲁棒性。(2) 当 $\alpha_1 = 0.8$, $\alpha_2 = 0.2$ 时, MMT-Net 在两组数据集上均得到最优的结果。

表 6 超参数 α_1 和 α_2 的消融实验结果

Table 6 Ablation experiment's results of hyper-parameter α_1, α_2

α_1	α_2	Inria				WHU			
		Precision	Recall	F1	IoU	Precision	Recall	F1	IoU
0.5	0.5	0.8822	0.8888	0.8855	0.7946	0.9553	0.9531	0.9542	0.9124
0.6	0.4	0.8889	0.8854	0.8871	0.7979	0.9580	0.9530	0.9555	0.9149
0.7	0.3	0.8844	0.8876	0.8860	0.7953	0.9551	0.9548	0.9549	0.9138
0.8	0.2	0.8901	0.8889	0.8894	0.8008	0.9581	0.9553	0.9567	0.9170
0.9	0.1	0.8895	0.8803	0.8849	0.7935	0.9567	0.9529	0.9548	0.9135

注: 加粗数字代表最优值。

3.3.3 模型复杂度分析

本节分析不同网络的计算复杂度。为此, 表 7 给出不同网络的参数量和浮点运算次数 FLOPs (Floating point operations)。同时, 为展示 LMIEB 分支的轻量级设计效果, 表 7 也给出所提出网络 MMT-Net 两个分支的参数量和 FLOPs。

表 7 不同方法的参数量和浮点运算次数

Table 7 The numbers of model parameters and floating point operations for different methods

方法	参数量/M	FLOPs/G
ENet(2016)	0.35	0.57
SegNet(2017)	29.44	40.14
MMB-Net(2022)	54.47	24.67
Refine-UNet(2023)	14.28	19.34
MAP-Net(2021)	4.08	37.78
MMT-Net	MBESB	28.54
	LMIEB	0.92
	整体	29.48

从表 7 可知: (1) 本文方法 MMT-Net 的参数量为 29.48 M, 与对比方法 SegNet 接近, 高于 ENet、Refine-UNet 和 MAP-Net, 低于 MMB-Net。本文方法 MMT-Net 的 FLOPs 为 64.85 G, 高于其他对比方法。(2) 所提出的 LMIEB 分支具有较小的参数量和计算量: LMIEB 的参数量为 0.92 M, 约占所提出网络 MMT-Net 整体的 1/32; LMIEB 的 FLOPs 为 2.34 G, 约占 MMT-Net 的 1/28。

4 结 论

针对遥感建筑物提取, 提出一种集成主体边缘分离和多尺度信息提取的双分支深度学习网络 MMT-Net。主体边缘分离分支 MBESB 基于解耦思想和光流估计技术实现对建筑物特征的主体和边缘分离, 从而增强对边缘特征的特征能力。轻量级多尺度信息提取分支 LMIEB 通过设计一种注意力增强多尺度卷积模块, 并将其应用到网络的多个层级, 来充分提取不同尺度的建筑物特征。此外, 利用 MBESB 分支生成的主体和边缘特征, 给出一种主体和边缘特征辅助增强的损失函数来优化网络训练过程, 提升网络的性能。

Inria 数据和 WHU 数据的实验结果表明: 所提出网络 MMT-Net 的建筑物提取结果在定性和定量两方面均优于 ENet、SegNet、MMB-Net、Refine-UNet 和 MAP-Net 这 5 种对比方法。

所提出的 MMT-Net 网络通过两个分支解决建筑物提取任务, 取得了较优的效果。然而 MMT-Net 的 MBESB 分支具有较大的参数量和计算量, 使得 MMT-Net 具有较高的计算复杂度。今后将尝试通过优化网络结构、利用分组卷积和深度可分离卷积来降低 MBESB 分支的参数量和计算复杂度。

参考文献(References)

Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep con-

- volitional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Cai J H and Chen Y M. 2021. MHA-Net: multipath hybrid attention network for building footprint extraction from high-resolution remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 5807-5817 [DOI: 10.1109/JSTARS.2021.3084805]
- Chen G J, Yue X Y, Zhu Y N and Fu Y P. 2023. Study on building extraction algorithm of remote sensing image based on multi-scale feature fusion. *Computer Science*, 50(9): 202-209 (陈国军, 岳雪燕, 朱燕宁, 付云鹏. 2023. 基于多尺度特征融合的遥感图像建筑物提取算法研究. *计算机科学*, 50(9): 202-209) [DOI: 10.11896/jsjcx.220800086]
- Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V, van der Smagt P, Cremers D and Brox T. 2015. FlowNet: Learning optical flow with convolutional networks//*Proceedings of the IEEE International Conference on Computer Vision*. Santiago: IEEE: 2758-2766 [DOI: 10.1109/ICCV.2015.316]
- He H, Wang S Y, Zhao Q, Lv Z F and Sun D W. 2021. Building extraction based on U-net and conditional random fields//*6th International Conference on Image, Vision and Computing*. Qingdao: IEEE: 273-277 [DOI: 10.1109/ICIVC52351.2021.9526925]
- Ji S P, Wei S Q and Lu M. 2019. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1): 574-586 [DOI: 10.1109/TGRS.2018.2858817]
- Jung H, Choi H S and Kang M. 2022. Boundary enhancement semantic segmentation for building extraction from remote sensed image. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-12 [DOI: 10.1109/TGRS.2021.3108781]
- Li E, Femiani J, Xu S B, Zhang X P and Wonka P. 2015. Robust rooftop extraction from visible band images using higher order CRF. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8): 4483-4495 [DOI: 10.1109/TGRS.2015.2400462]
- Li J Y, Huang X, Tu L L, Zhang T and Wang L G. 2022. A review of building detection from very high resolution optical remote sensing images. *GIScience and Remote Sensing*, 59(1): 1199-1225 [DOI: 10.1080/15481603.2022.2101727]
- Li X T, Li X, Zhang L, Cheng G L, Shi J P, Lin Z C, Tan S H and Tong Y H. 2020. Improving semantic segmentation via decoupled body and edge supervision//*16th European Conference on Computer Vision*. Glasgow: Springer: 435-452 [DOI: 10.1007/978-3-030-58520-4_26]
- Li Z, Sui Z W, Fu Q Y, Zheng J J and Bu T. 2023. High-resolution remote sensing extraction of urban buildings based on morphological sequences and multi-source a priori information. *National Remote Sensing Bulletin*, 27(4): 998-1008 (李治, 隋正伟, 傅俏燕, 郑珊珊, 卜桐. 2023. 基于形态学序列和多源先验信息的城市建筑物高分遥感提取. *遥感学报*, 27(4): 998-1008) [DOI: 10.11834/jrs.20221077]
- Liu Y Y, Chen D Y, Ma A L, Zhong Y F, Fang F and Xu K. 2021. Multiscale U-shaped CNN building instance extraction framework with edge constraint for high-spatial-resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7): 6106-6120 [DOI: 10.1109/TGRS.2020.3022410]
- Luo L, Li P P and Yan X S. 2021. Deep learning-based building extraction from remote sensing images: a comprehensive review. *Energies*, 14(23): 7982 [DOI: 10.3390/en14237982]
- Maggiori E, Tarabalka Y, Charpiat G and Alliez P. 2017. Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark//*IEEE International Geoscience and Remote Sensing Symposium*. Fort Worth: IEEE: 3226-3229 [DOI: 10.1109/IGARSS.2017.8127684]
- Paszke A, Chaurasia A, Kim S and Culurciello E. 2016. ENet: a deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv: 1606.02147* [DOI: 10.48550/arXiv.1606.02147]
- Qiu W Y, Gu L J, Gao F and Jiang T. 2023. Building extraction from very high-resolution remote sensing images using refine-UNet. *IEEE Geoscience and Remote Sensing Letters*, 20: 6002905 [DOI: 10.1109/LGRS.2023.3243609]
- Sheikh M A A, Maity T and Kole A. 2022. IRU-net: an efficient end-to-end network for automatic building extraction from remote sensing images. *IEEE Access*, 10: 37811-37828 [DOI: 10.1109/ACCESS.2022.3164401]
- Sun Y, Zhang X C, Zhao X Y and Xin Q C. 2018. Extracting building boundaries from high resolution optical images and LIDAR data by integrating the convolutional neural network and the active contour model. *Remote Sensing*, 10(9): 1459 [DOI: 10.3390/rs10091459]
- Tian Q L, Zhao Y J, Li Y, Chen J, Chen X J and Qin K. 2022. Multiscale building extraction with refined attention pyramid networks. *IEEE Geoscience and Remote Sensing Letters*, 19: 8011305 [DOI: 10.1109/LGRS.2021.3075436]
- Yang G, Zhang Q and Zhang G X. 2020. EANet: edge-aware network for the extraction of buildings from aerial images. *Remote Sensing*, 12(13): 2161 [DOI: 10.3390/rs12132161]
- Zhang H, Zheng X C, Zheng N S and Shi W Z. 2022. A multiscale and multipath network with boundary enhancement for building footprint extraction from remotely sensed imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 8856-8869 [DOI: 10.1109/JSTARS.2022.3214485]
- 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: 5515714 [DOI: 10.1109/TGRS.2023.3282926]
- Zhao Y Y, Zheng G Z, Xu Z Y, Qiu Z H and Chen Z X. 2022. Multiscale feature weighted-aggregating and boundary enhancement network for semantic segmentation of high-resolution remote

sensing images. IEEE Journal of Selected Topics in Applied Earth
Observations and Remote Sensing, 15: 8118-8130 [DOI: 10.1109/
JSTARS.2022.3205609]
Zhu Q, Liao C, Hu H, Mei X M and Li H F. 2021. MAP-Net: multiple

attending path neural network for building footprint extraction
from remote sensed imagery. IEEE Transactions on Geoscience
and Remote Sensing, 59(7): 6169-6181 [DOI: 10.1109/TGRS.
2020.3026051]

A two-branch building extraction network integrating main body and edge separation and multi-scale information extraction

SONG Baogui^{1,2,3}, SHAO Pan^{1,2,3}, SHAO Wen^{1,2,3}, ZHANG Xiaodong⁴, DONG Ting^{1,2,3}

1. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang 443002, China;

2. College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China;

3. Advanced Computing Center, China Three Gorges University, Yichang 443002, China;

4. State Key Laboratory of Remote Sensing Information Engineering for Surveying and Mapping, Wuhan University, Wuhan 430079, China

Abstract: Building extraction in Remote Sensing (RS) images is a key research focus in RS, which plays a vital role in various tasks, such as urban planning, illegal building detection, and natural disaster assessment. The rapid development of deep learning technology has facilitated its use in building extraction in RS images, which achieved significant extraction effect. However, building extraction in RS images has two main challenges: one is the difference in scales and shapes of buildings in RS images, and the other is the difficulty in accurately extracting building boundaries. This study proposes a two-branch building extraction network, which integrates main body and edge separation and multiscale information extraction, to address the aforementioned challenges. Both branches employ the U-Net architecture. First, a Main Body and Edge Separation Branch (MBESB) is designed for feature decomposition based on the decoupling idea and optical flow estimation technique. MBESB generates the main body and edge features of buildings separately, which enhances the ability of representing building boundaries. Then, a Lightweight Multiscale Information Extraction Branch (LMIEB) is constructed based on dilated convolution, depth-separable convolution, and attention mechanism to fully extract the differently scaled features of buildings. LMIEB introduces an Attention-enhanced Multi-scale Convolutional Module (AMCM) and integrates this module into the last three stages of the U-shaped network to harvest multi-scale information from feature maps at different levels. Finally, a loss function enhanced by main body and edge features is presented with the help of the features generated by MBESB and LMIEB to improve the training process of the building extraction network. The loss function consists of three components: a global loss based on overall building features, a main body loss focused on the building's body features, and an edge loss emphasizing boundary details. Experiments were conducted using two datasets for public building extraction, namely, the Inria and WHU datasets, to evaluate the performance of the proposed MMT-Net method. Moreover, to further validate the effectiveness of the proposed method, extensive comparative experiments were carried out on subregions Tyrol, Austin, and Chicago of the Inria dataset. Five deep learning methods (ENet, SegNet, MMB-Net, Refine-UNet, and MAP-Net) were used as the comparative methods. Quantitative analysis of building extraction results was conducted in terms of four evaluation metrics, namely, precision, recall, F1, and IoU. For the Inria and WHU datasets, the F1/IoU values of the proposed MMT-Net are 0.8894/0.8008 and 0.9567/0.9170, respectively, which are superior to those of the five comparative methods. Additionally, visualization results are shown on both Inria and WHU datasets. The findings indicate that the proposed method outperforms other five comparative methods in terms of both missed building detections and false building detections. Experimental results on two commonly used datasets for public building extraction show that the proposed building extraction network is feasible and effective. In addition, the results of the ablation experiments indicate that MBESB, LMIEB, and the loss function with auxiliary enhancement of the main body and edge features proposed in this work can effectively enhance the performance of building extraction.

Key words: remote sensing image, building extraction, deep learning, U-Net, main body and edge separation, two-branch, multi-scale, lightweight

Supported by National Natural Science Foundation of China (No. 41901341); Nature Science Foundation of Hubei Province (No. 2024AFB867)