



DOI:10.12404/j.issn.1671-1815.2404716

引用格式:何锐利,乐伟鹏,俞友,等.坐标注意力及卷积增强的全尺度融合建筑物提取网络[J].科学技术与工程,2025,25(18):7485-7492.

He Ruili, Le Weipeng, Yu You, et al. Full scale fusion building extraction network with coordinate attention and convolution enhancement [J]. Science Technology and Engineering, 2025, 25(18): 7485-7492.

天文学、地球科学

坐标注意力及卷积增强的全尺度融合 建筑物提取网络

何锐利^{1,2}, 乐伟鹏³, 俞友^{1,2}, 黄亮^{3*}

(1. 湖南省地质灾害调查监测所, 长沙 410004; 2. 湖南省勘测设计院有限公司, 长沙 410004;
3. 昆明理工大学国土资源工程学院, 昆明 650093)

摘要 建筑物作为人类生产活动的重要载体,准确快速地提取建筑物可在自然资源管理领域发挥重要作用。基于卷积神经网络(convolutional neural network, CNN)在遥感影像建筑物提取方面取得了重大进展,但构建的网络模型在特征提取和特征融合方面仍有待优化。因此,提出了一种坐标注意力及卷积增强的全尺度融合建筑物提取网络(coordinate attention and convolution enhanced full-scale fusion building extraction network, CCFNet)。所构建的模型由坐标注意力及卷积增强的残差编码器和全尺度融合解码器组成。编码器使用坐标注意力构建通道间的依赖关系并捕获的全局信息,其使用的非对称卷积增强地物边缘特征提取,并对旋转、翻转扭曲及纵横比不均匀的地物有更强的鲁棒性。解码器使用的全尺度融合方法则有助于建筑物的重建。在中国典型城市建筑物实例数据集实验结果表明,相比于其他建筑物提取网络,本文构建的CCFNet模型在Accuracy、 F_1 、IOU和MIOU共4种分割评价指标分别取得了93.84%、84.08%、72.53%和82.59%的最优实验精度。结果表明,该模型能够有效地提取建筑物区域。

关键词 坐标注意力;全尺度融合;建筑物提取;非对称卷积

中图法分类号 P237;

文献标志码 A

Full Scale Fusion Building Extraction Network with Coordinate Attention and Convolution Enhancement

HE Rui-li^{1,2}, LE Wei-peng³, YU You^{1,2}, HUANG Liang^{3*}

(1. Hunan Geological Hazards Investigation and Monitoring Institute, Changsha 410004, China;

2. Hunan Survey and Design Institute Co., Ltd., Changsha 410004, China;

3. School of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China)

[Abstract] Buildings are important carriers of human production activities. Accurate and fast extraction of building areas can play an important role in the field of natural resource management. Although significant progress has been made in building extraction from remote sensing images based on CNN (convolutional neural network), the constructed network model still needs to be optimized in feature extraction and feature fusion. Therefore, a coordinate attention and CCFNet (convolution enhanced full-scale fusion building extraction network) was proposed. The constructed model consists of a residual encoder enhanced by coordinate attention and convolution and a full-scale fusion decoder. Coordinate attention was used in the encoder to build inter-channel dependencies and capture global information. The asymmetric convolution was used to enhance the edge feature extraction of ground objects, and it is more robust to rotation, flip distortion and uneven aspect ratio of ground objects. The full-scale fusion method used in the decoder helps to reconstruct the buildings. The experimental results on the dataset of typical Chinese city buildings show that compared with other building extraction networks, The CCFNet model constructed in this paper achieves the best experimental Accuracy of 93.84%, 84.08%, 72.53% and 82.59% in the four segmentation evaluation indicators of accuracy, F_1 , IOU and MIOU, respectively. Experimental results show that the model can effectively extract building regions.

收稿日期: 2024-06-24 修订日期: 2025-03-19

基金项目: 湖南省自然科学基金(2023JJ60561); 云南省基础研究计划(202201AT070164); 兴滇英才支持计划(KKRD202221036)

第一作者: 何锐利(1982—), 男, 汉族, 湖南邵阳人, 硕士, 高级工程师。研究方向: 自然资源调查监测和地质灾害的识别监测预警。E-mail: 276105885@qq.com。

* 通信作者: 黄亮(1985—), 男, 汉族, 江西进贤人, 博士, 副教授。研究方向: 变化检测、语义分割。E-mail: kmhuangliang@163.com。

投稿网址: www.stae.com.cn

[Keywords] coordinate attention; full scale fusion; building extraction; asymmetric convolution

建筑物作为城市体系中的基本实体,是人类生产生活的主要载体。利用遥感影像进行建筑物提取,即识别影像中的建筑物和非建筑物像素,在城市规划、人口统计、经济评估和输电线沿线隐患区识别^[1]等广泛的应用场景中发挥着至关重要的作用。以城市规划与管理为例,通过高效、准确的建筑物提取算法,可以快速获取城市中建筑物的分布、类型等信息,为城市管理和决策提供科学依据。在灾害应急评估中,快速、准确地从遥感影像中提取建筑物对于灾害发生后的快速响应和救援行动至关重要,能够帮助评估灾害影响范围和程度,指导救援资源的分配。传统的建筑提取方法通常提取人工设计特征,并应用传统的机器学习方法来识别建筑物。樊华等^[2]通过对差分图像采用 W_v -Canny 边缘检测方法获得差异区域重要地物的边缘信息,进而识别地物变化。然而,经验设计的特征限制了传统方法的泛化能力。深度学习卷积神经网络(convolutional neural network, CNN)^[3]作为机器学习新范式,是一种自动捕获非线性和分层特征的强大方法。近年来,大多数遥感影像分割方法都使用深度卷积神经网络,其性能普遍优于传统方法,并在具有挑战性的数据集中取得了最先进的结果。从遥感图像中提取建筑物类似于计算机视觉社区中开发的语义分割任务,旨在将图像的整个像素标记为建筑物类或非建筑物类。

目前开展遥感图像建筑物提取研究的全卷积神经网络,是通过构造编码器-解码器结构,解决粗分辨率分割问题。如 UNet^[4]和 SegNet^[5]等语义分割方法逐步恢复了提取特征的空间分辨率,以实现精细分辨率的特征表示。例如, Kang 等^[6]设计了一种基于密集空间金字塔模块,提取建筑物的多尺度特征; Zhu 等^[7]提出了一种新的多参与路径神经网络,用于精确提取多尺度建筑物覆盖区和精确边界;何直蒙等^[8]提出的建筑物提取网络在 UNet 中使用空洞卷积增加感受野; Chen 等^[9]提出了一种密集残差神经网络,将密集连接的卷积神经网络与残差网络结构相结合,可以高效提取建筑特征。虽然这些基于 CNN 的建筑物提取方法取得了巨大的进展,但在建筑物特征提取和特征融合方面仍有待优化。比如,固定大小的卷积核会导致局部感受野的问题,虽然可以堆叠更多的卷积层来获得图像的全局信息,但这种方式可能会导致训练的模型更复杂。采用注意力机制能够有效地捕捉整体关系,从全局视角关注重要位置,获取更多有用信息^[10]。在

建筑物提取网络中使用注意力机制,神经网络能够自动地学习并选择性地关注输入中的重要信息,提高模型的性能和泛化能力。例如, Liu 等^[11]构建的建筑物提取网络中利用压缩和激活(squeeze and excitation, SE)注意力机制来提高建筑物的特征权重,从而提高建筑物的提取性能。Cai 等^[12]提出了一种多路径混合注意网络用于建筑自动提取并使用通道注意模块捕获通道的全局关系,从而提高分割性能。曲海成等^[13]提出了融合混合注意力机制与多尺度特征增强的高分辨率建筑物语义分割网络,使用了通道空间注意机制(convolutional block attention module, CBAM),增强高层特征的表示能力。SE 注意力机制^[14]只考虑编码通道间信息,而忽略了位置信息的重要性,而位置信息对于视觉任务中捕获目标结构至关重要。CBAM 注意力机制^[15]通过降低输入张量的通道维数,然后使用卷积计算空间注意力来利用位置信息,然而卷积只能捕获局部关系,而无法建模视觉任务所需的全局信息。相比于 SE 和 CBAM 通道注意力机制,坐标注意力能够构建通道间的依赖关系并捕获的全局信息。

因此,现引入坐标注意力(coordinate attention, CA)^[16]嵌入残差卷积块中,增强网络特征的表达能力。与此同时,受水平和垂直卷积良好的轮廓保持优势及其对旋转或翻转扭曲数据的强大鲁棒性的启发^[17]。在构建的编码器残差卷积块中使用非对称卷积代替 3×3 卷积,以增强地物边缘特征提取,并对旋转、翻转扭曲及纵横比不均匀的地物有更强的鲁棒性。在解码重建建筑物阶段,相比于采用单一尺度跳跃连接的 Unet 建筑物提取基准网络^[8]、具有嵌套和密集连接 Unet + 建筑物提取基准网络^[18]。本文研究中使用的 Unet3 +^[19]基准架构,其跳跃连接采用编码器和解码器之间的连接以及解码器子网之间的内部连接,以全尺度捕获细粒度细节和粗粒度语义,有助于建筑物的重建。

1 网络架构

1.1 总体架构

本文所提的网络模型如图 1 所示。模型主要由两部分组成,分别为特征提取编码器和特征融合解码器。采用的编码器为多个坐标注意力及卷积增强的残差卷积块组成,其具体网络结构如图 1 左边虚线框所示。采用的解码器融合方式如图 1 右边虚线框所示。编码器总共由 5 层残差卷积块组成,每层通过下采样操作降低图像的分辨率。其每一层

分辨率分别为原始图像的 1、1/2、1/4、1/8 和 1/16 大小。原始图像输入的通道数为 3,大小为 $H \times W$ 。经过卷积、下采样操作之后,每一层的输入特征其通道数分别为 64、128、256、512 和 1 024。首先遥感图像输入建筑物提取网络中,提取多尺度特征 E1、E2、E3、E4 和 E5。其次将多尺度特征进行全尺度特征融合得到解码器特征 D4、D3、D2 和 D1。最后使用混合损失函数对模型进行训练。

1.2 特征提取编码器

在建筑物提取任务中,检测对象的形状、大小和方向呈现出不一样变化。基于 CNN 的建筑物提取网络,其卷积基本单元大多为 3×3 卷积,但 3×3 卷积难以提取例如竖直或水平的边缘和纹理等非对

称特征。而 1×3 或 3×1 这类非对称卷积则能够更好地捕捉图像中的非对称特征。所以本文在特征提取编码器用非对称卷积块代替 3×3 卷积,使用的非对称卷积块由 1×3 、 3×1 、 3×3 构成 3 个卷积单元组成,且每个卷积单元都经过卷积、批量归一化 (batch normalization, BN)、Rule 激活函数过程。所构成的 3 个卷积单元一起共同作为一个非对称卷积块对特征图进行卷积运行。由于二维卷积的可加性,其非对称卷积块相当于在 3×3 卷积的基础上,叠加了 1×3 或 3×1 卷积,以提取竖直或水平的边缘和纹理非对称特征,并对旋转、翻转扭曲及纵横比不均匀的地物有更强的鲁棒性。其过程如图 2 所示。

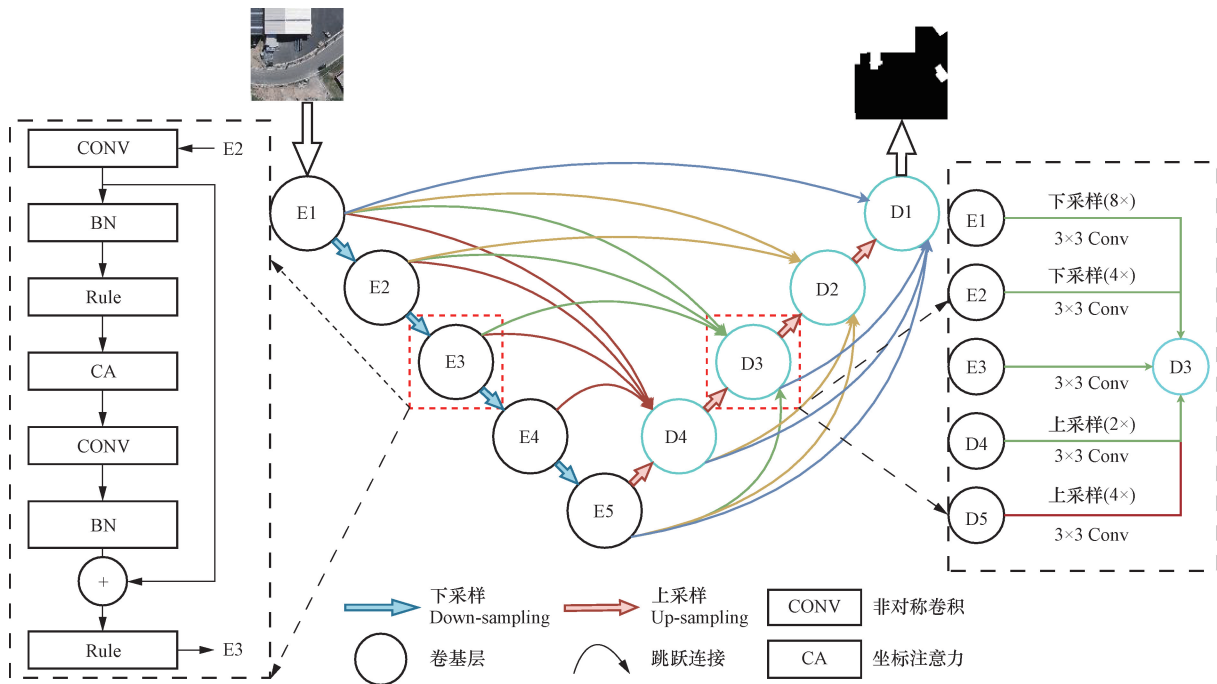


图 1 总体网络架构

Fig. 1 Overall network architecture

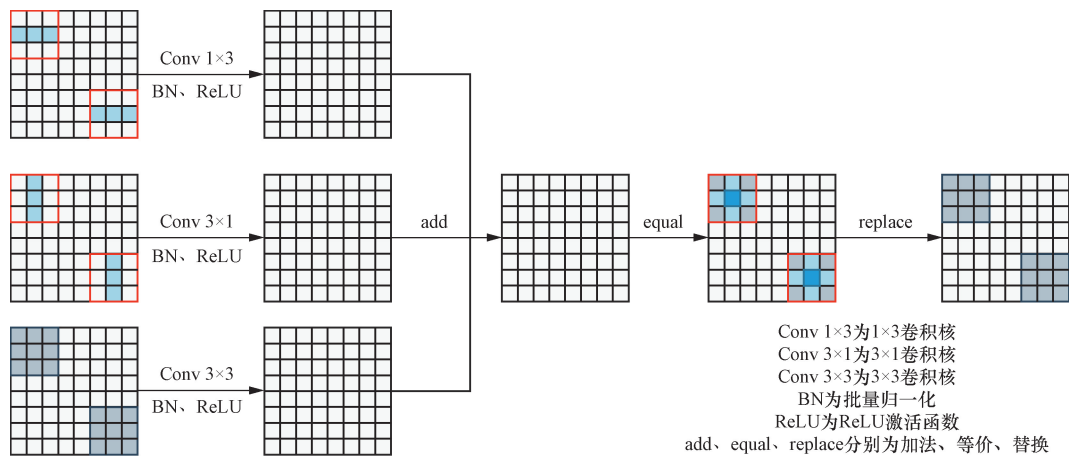


图 2 非对称卷积块

Fig. 2 The block of asymmetric convolutional

在非对称卷积块的基础上,设计了一种坐标注意力及非对称卷积增强的残差卷积块提取多尺度特征,增强多尺度特征的提取能力。与普通的残差卷积块相比,本文提出的坐标注意力及非对称增强的残差卷积块可以提供更多的模式和方向的特征学习,这使得模型可以学习到更丰富和多样的特征,从而提高建筑物网络模型的性能。本文提出的残差卷积块如图1左边虚线框所示。该模块由2个非对称卷积块、2个批量归一化、2个激活函数、1个坐标注意力机制和1个跳跃连接组成。首先输入的特征图使用非对称卷积提取非对称特征,其次在经过批量归一化、Rule 激活函数和坐标注意力机制,然后在经过一层非对称卷积块、批量归一化以提取更深层次特征,最后将上一个非对称卷积块提取特征与下一个非对称卷积块提取特征进行逐元素相加输出特征图,经过Rule 激活函数输出结果特征图。经过5次残差卷积来获取每一层级的特征图,其中每一层级之间经过一次下采样过程,以构建特征提取编码器。其编码器每一层运算过程公式为

$$\text{Conv}_{\text{ADC}}(f_i^n) = R\{B_N[\text{Conv}_{3 \times 1}(f_i^n)]\} + R\{B_N[\text{Conv}_{1 \times 3}(f_i^n)]\} + R\{B_N[\text{Conv}_{3 \times 3}(f_i^n)]\} \quad (1)$$

$$\text{output} = R[\text{Conv}_{\text{ADC}}(f_i^n) + B_N \times \{\text{Conv}_{\text{ADC}}[\text{CA}(R\{B_N[\text{Conv}_{\text{ADC}}(f_i^n)]\})\}]] \quad (2)$$

式中: $\text{Conv}_{3 \times 1}$ 为 3×1 卷积核; $\text{Conv}_{1 \times 3}$ 为 1×3 卷积核; $\text{Conv}_{3 \times 3}$ 为 3×3 卷积核; Conv_{ADC} 为非对称卷积块; B_N 和 R 含义为BN和Rule;CA为坐标注意力机制; f_i^n 为第 n 层编码输入特征,其中 $n=1,2,3,4,5$; $i=0,1$ 。

在本文构建的残差卷积块中,嵌入了CA,旨在增强网络的表达能力。除此之外由于CA的轻量化,不会对网络计算带来沉重的负担。CA是通过精确的位置信息对通道关系和长期依赖性进行编码。具体分为坐标信息嵌入和坐标注意生成2部分。给定输入特征张量 $\mathbf{X} = [X_1, X_2, \dots, X_C]$,使用大小为 $(H,1)$ 或 $(1,W)$ 的池化核分别沿着水平坐标和垂直坐标对每个通道进行编码。因此,高度为 H 的第 C 通道的输出和宽度为 W 的第 C 通道的输出的公式为

$$Z_C^H(H) = \frac{1}{W} \sum_{0 \leq i < W} X_C(H, i) \quad (3)$$

$$Z_C^W(W) = \frac{1}{H} \sum_{0 \leq j < H} X_C(j, W) \quad (4)$$

式中: i 和 j 为输入特征图的空间维度(高度和宽度)上的位置索引。

上述2种变换分别沿两个空间方向聚合特征,得到一对方向感知的特征图。通过信息嵌入中的变换后,该部分将上面的变换进行通道拼接操作,然后使用 1×1 卷积变换函数 K_1 对其进行变换操作。

$$f = \delta[K_1([Z^H, Z^W])] \quad (5)$$

$$G^H = \sigma[K_H(f^H)] \quad (6)$$

$$G^W = \sigma[K_W(f^W)] \quad (7)$$

式中: $[\cdot]$ 为通道拼接操作; δ 为非线性激活函数; f 为对空间信息在水平方向和垂直方向进行编码的中间特征映射,然后沿着空间维数将 f 分解为2个单独的张量 f^H 和 f^W 。 K_H 和 K_W 为另外2个 1×1 卷积变换; σ 为sigmoid激活函数。

最后输出的计算公式为

$$\text{CA}_C(i, j) = X_C(i, j) G_C^H(i) G_C^W(j) \quad (8)$$

式(8)中: $\text{CA}_C(i, j)$ 为输出特征; $X_C(i, j)$ 为输入特征在 (i, j) 处通道 C 的原始特征值; $G_C^H(i)$ 为通道 C 在高度位置 i 处的注意力系数; $G_C^W(j)$ 为通道 C 在宽度位置 j 的注意力系数。

1.3 全尺度特征融合解码器

融合高层次语义信息和低层次细节信息,以及在不同尺度上的信息,可以有效重建建筑物区域。所以设计一种全尺度特征融合网络至关重要。因此,本文构建了全尺度特征融合解码器,聚合了多尺度编码器特征图及解码器特征图以获取解码器每一层级融合特征图,捕获编码器不同层级特征图的细节信息和语义信息。其构建的建筑物提取网络解码器如图1右边虚框所示。以下说明D3特征图融合过程。首先将E1和E2下采样到E3空间分辨率大小,而D4和D5上采样到E3空间分辨率大小。然后将E1、E2、E3、D4和D5卷积为64通道数。最后将E1、E2、E3、D4和D5卷积后的通道数进行沿通道方向进行拼接,得到D3特征图。

1.4 混合损失函数

本文定义的混合损失函数由BCE^[20]、IOU^[21]和SSIM^[22]损失函数组成。其中IOU损失是基于像素的损失函数,它不考虑像素邻域值,并且前景和背景的权重值相等。SSIM损失是基于补丁的损失函数,它考虑每个像素的局部邻域。它为边界分配了更高的权重,即边界上的预测概率和前景的其余部分相同,边界周围的损失也更高。混合损失函数计算公式分别为

$$L_{\text{bce}} = - \sum_{r=1}^H \sum_{c=1}^W \{G(r, c) \ln[S(r, c)] + [1 - G(r, c)] \ln[1 - S(r, c)]\} \quad (9)$$

$$L_{\text{ssim}} = 1 - \frac{(2u_x u_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (10)$$

$$L_{iou} = 1 - \frac{\sum_{r=1}^H \sum_{c=1}^W S(r,c)G(r,c)}{\sum_{r=1}^H \sum_{c=1}^W [S(r,c) + G(r,c) - S(r,c)G(r,c)]} \quad (11)$$

$$L_{total} = L_{bce} + L_{ssim} + L_{iou} \quad (12)$$

式中: r, c 分别为像素的横、纵坐标; $G(r, c) \in \{0, 1\}$, 为像素 (r, c) 的真值; $S(r, c)$ 为目标的预测概率; $x = \{x_j: j = 1, 2, \dots, N^2\}$ 、 $y = \{y_j: j = 1, 2, \dots, N^2\}$, 分别为从预测概率图 S 和二值地面真值 G 中裁剪出的两个对应的切片(大小为 $N \times N$) 的像素值; H 和 W 分别为高度和宽度; μ_x, μ_y 和 σ_x, σ_y 分别为 x, y 的均值和标准差; σ_{xy} 为协方差, 其中 $C_1 = 0.012, C_2 = 0.032$ 。

2 分析与讨论

首先介绍了中国典型城市建筑物实例数据集^[23]。接下来,介绍了模型一些参数设置及评价结果所使用的各项指标。最后定性和定量的评价了对比实验和消融实验结果。

2.1 数据集

中国典型城市建筑物实例数据集选取了北京、上海、深圳、武汉城市中心作为数据采集点,其原始数据为谷歌卫星影像,包括正射影像和非正射影像区域,空间分辨率为 0.29 m。其数据集总共有 7 260 个样本区域。每个样本大小为 500 像素 \times 500 像素。本文研究中将数据集划分为训练集 5 985 张图像、测试集 1 275 张图像。其示例图如图 3 所示。

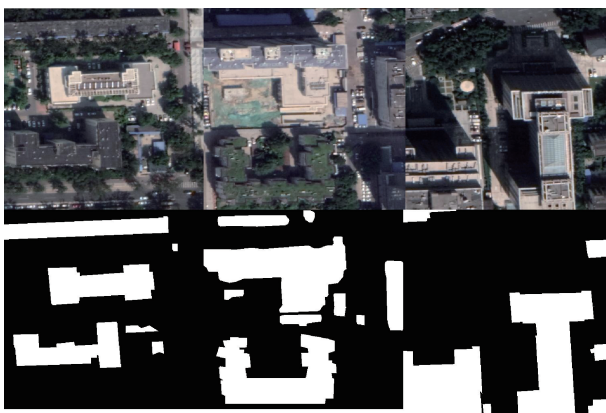


图 3 中国典型城市建筑物实例数据集示例图

Fig. 3 Example data set of typical city buildings in China

2.2 实验细节

本文模型使用的深度学习框架为 pytorch。模型使用的 GPU 为 NVIDIA GeForce RTX 3090, 内存 64 GB, 显存 24 GB, 模型优化器为 Adam。迭代周期

为 50 次。模型学习率为 0.001, 批量大小为 2。模型接受 3 通道大小的图片。模型使用总体精度 (accuracy, A_{cc})、 F_1 分数 (F_1)、交并比 (I_{OU}) 和平均交并比 (M_{IOU}) 4 种精度指标评价实验结果。

2.3 实验结果

为了验证本文模型的性能和有效性,本文模型与流行的几种的语义分割模型进行对比。采用的对比模型分别为 Unet、Unet++ 和 SEUnet。表 1 为 4 个模型在测试数据集上的平均分割性能精度。从表 1 中可知,CCFNet 模型相比于 Unet、SEUnet 和 Unet++ 在 4 种评价指标上取得了最佳的精度。 A_{cc} 、 F_1 、 I_{OU} 和 M_{IOU} 精度分别为 93.84%、84.08%、72.53% 和 82.59%。相比于 Unet++ 在 4 个评价指标上分别提高了 0.41%、2.49%、3.62% 和 1.98%。

图 4 为对比实验可视化结果。在密集建筑物场景情况下,各个分割模型建筑物提取的实验效果。图 4(c)~图 4(f) 分别 Unet、SEUnet、Unet++ 和 CCFNet。图 4(c) 中可以看出,Unet 模型的实验结果基本可以检测出大部分建筑区域,但其检测的边缘区域呈现出不连续性。从图 4(d) 中可以看出,其实验结果相比于 Unet 模型结果,建筑物边界更加的完整且也能够检测出细微的建筑物区域。如红色框中所示,Unet++ 相比于 Unet 和 SEUnet 其检测结果呈现出更多漏检和误检现象。但相比于以上 3 种模型,CCFNet 的建筑物检测结果效果更好,基本能够检测出大部分建筑物区域,具有更少的漏检和误检问题。从红色框也可以看出,CCFNet 模型提取的建筑物边缘区域更加了完整和连续。从模型上分析,相比于 Unet 等网络,CCFNet 采用的 Unet3+ 解码器采用全尺度融合结合深浅层信息,可以更好地重建建筑区域,且使用的非对称卷积增强了边缘变化区域,而 CA 注意力机制则增强了通道间的信息,能够捕获全局的信息。

表 2 为消融实验结果。其中 Base 为不使用非对称卷积 (asymmetric convolution, AC) 和坐标注意力 (coordinate attention, CA) 的模型实验结果。其在 4 种评价指标 A_{cc} 、 F_1 、 I_{OU} 和 M_{IOU} 上分别取得 92.90%、81.27%、68.45% 和 80.03% 的实验结果。而添加了 AC 卷积后,相比于 Base 模型,其 A_{cc} 、 F_1 、 I_{OU}

表 1 对比实验评价结果

对比实验模型	$A_{cc}/\%$	$F_1/\%$	$I_{OU}/\%$	$M_{IOU}/\%$
Unet	92.20	79.47	65.93	78.38
SEUnet	93.14	80.03	66.71	79.38
Unet++	93.43	81.59	68.91	80.61
CCFNet	93.84	84.08	72.53	82.59

注:加粗数值表示最优结果。

表2 消融实验结果

Table 2 Results of ablation experiments

消融实验模型	$A_{cc}/\%$	$F_1/\%$	$I_{ou}/\%$	$M_{100}/\%$
Base	92.90	81.27	68.45	80.03
Base + AC	93.61	82.90	70.79	81.62
Base + AC + CA	93.84	84.08	72.53	82.59

注:加粗数值表示最优结果。

和 M_{100} 分别取得 0.71%、1.63%、2.34% 和 1.59% 的精度提升。而继续在 Base 添加 CA 后,其实验精度相比于 Base 模型,其 A_{cc} 、 F_1 、 I_{OU} 和 M_{100} 分别取得 0.94%、2.81%、4.08% 和 2.56% 的精度提升。实验表明,使用了 AC 和 CA 之后,模型的精度取得了一定的提升。图 5 为消融实验可视化结果。从红色

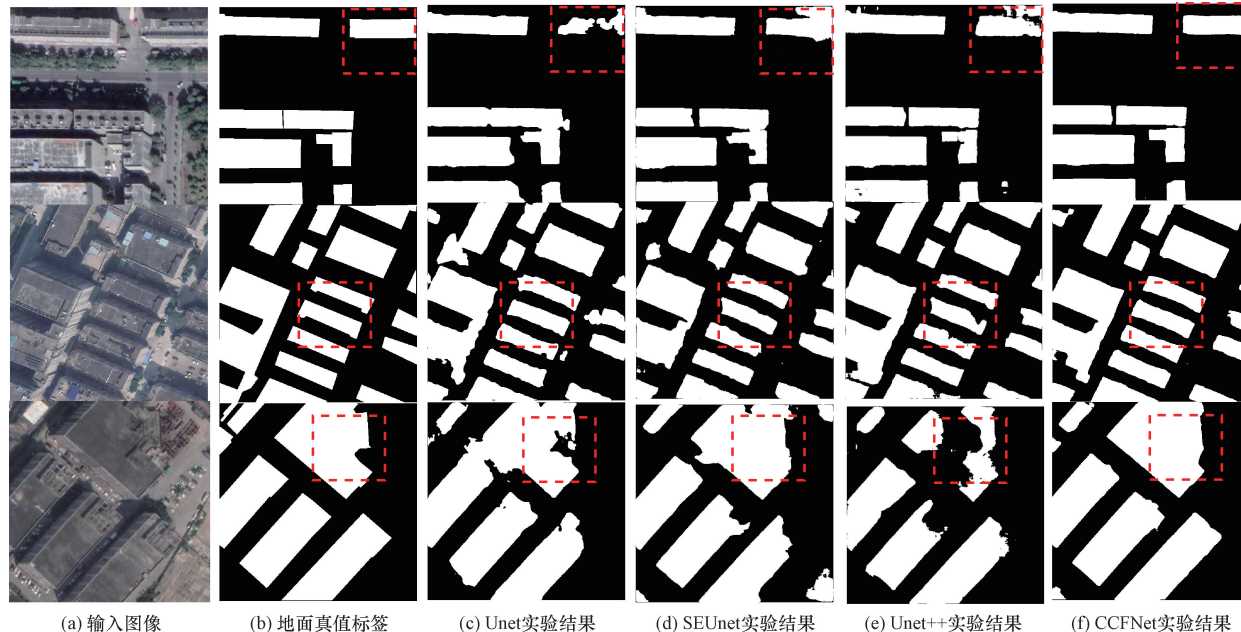


图4 对比实验可视化结果

Fig. 4 The visualization results of comparison experimental

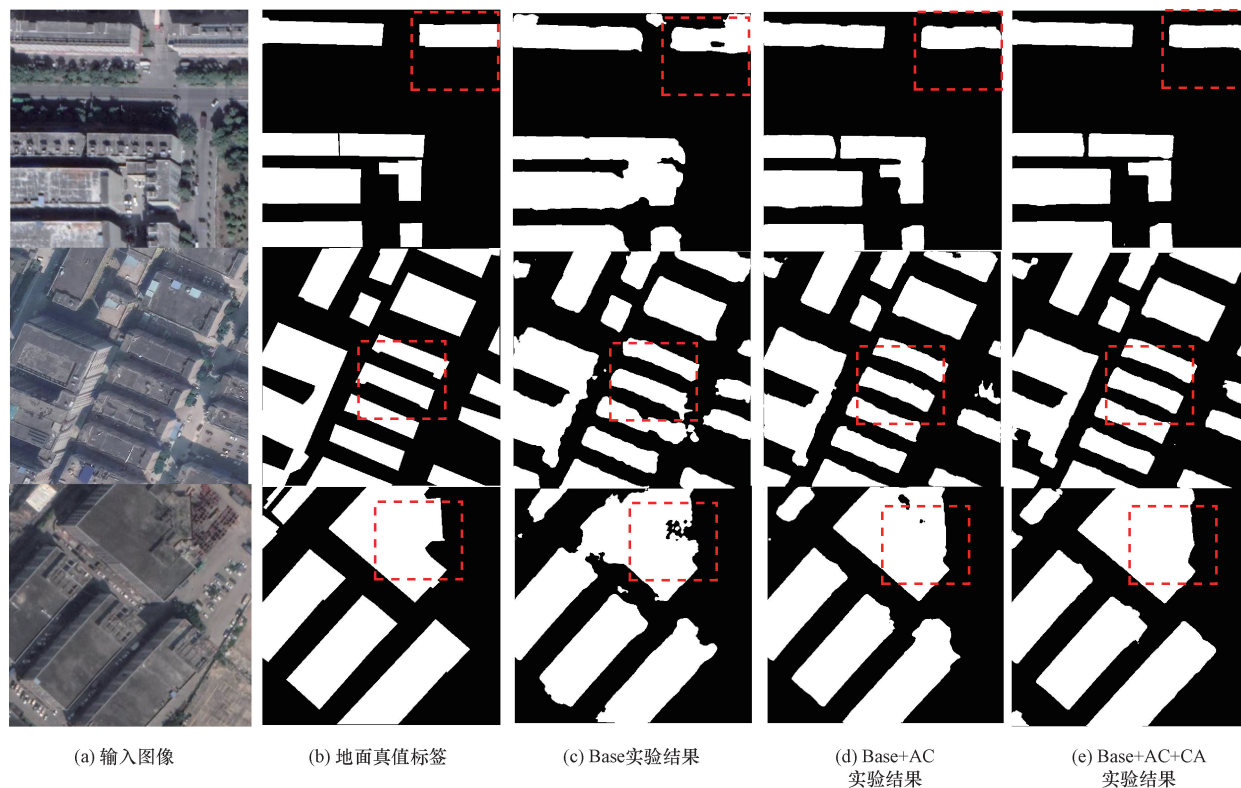


图5 消融实验结果图

Fig. 5 The results of ablation experiment

框可以看出,使用了 AC 之后,建筑物边缘区域得到一定的改善。这得益于非对称卷积可以增强地物边缘特征提取,并对旋转、翻转扭曲及纵横比不均匀的地物有更强的鲁棒性。在模型中添加 CA 模块之后,其实验结果效果得到了进一步的提升。

3 结论

为增强建筑物提取网络的性能,提出了一种坐标注意力及卷积增强的全尺度融合建筑物提取网络(CCFNet)。所构建的模型由坐标注意力及卷积增强的残差编码器和全尺度融合解码器组成。通过对比实验和消融实验,表明了 CCFNet 模型的有效性,也证实了坐标注意力和非对称卷积可以有效地增强模型的实验精度,并可以完整地重建建筑物区域及处理建筑物边界信息。在中国典型城市建筑物实例数据集实验结果表明,相比于其他建筑物提取网络,CCFNet 取得了最优的精度评价分数。但本文提出的 CCFNet 虽然能够提取绝大部分建筑物区域,但模型仍存在一些问题,如模型的训练时间长、模型训练数据量大的问题。未来研究需进一步关注模型的轻量化工作,减少模型的参数量和运行时间。

参 考 文 献

- [1] 周恺,张睿哲,谭磊,等. 基于卫星遥感的输电沿线建筑物变化检测[J]. 科学技术与工程, 2023, 23(35): 15100-15108.
Zhou Kai, Zhang Ruizhe, Tan Lei, et al. Building change detection along transmission line based on satellite remote sensing[J]. Science Technology and Engineering, 2023, 23(35): 15100-15108.
- [2] 樊华,王文旭,孙杰,等. 基于遥感影像的重要地物的变化检测和标注[J]. 科学技术与工程, 2024, 24(9): 3586-3595.
Fan Hua, Wang Wenxu, Sun Jie, et al. Change detection and annotation of important ground objects based on remote sensing images [J]. Science Technology and Engineering, 2024, 24(9): 3586-3595.
- [3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [4] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[J]. Lecture Notes in Computer Science, 2015, 9351: 234-241.
- [5] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [6] Kang W, Xiang Y, Wang F, et al. EU-Net: an efficient fully convolutional network for building extraction from optical remote sensing images [J]. Remote Sensing, 2019, 11(23). DOI: 10.3390/rs11232813.
- [7] Zhu Q, Liao C, Hu H, et al. MAP-Net: multiple attending path neural network for building footprint extraction from remote sensed imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(7): 6169-6181.
- [8] 何直蒙,丁海勇,安炳琪. 高分辨率遥感影像建筑物提取的空洞卷积 E-U-Net 算法[J]. 测绘学报, 2022, 51(3): 457-467.
He Zhimeng, Ding Haiyong, An Bingqi. E-Unet: a atrous convolution-based neural network for building extraction from high-resolution remote sensing images [J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(3): 457-467.
- [9] Chen M, Wu J, Liu L, et al. DR-Net: an improved network for building extraction from high resolution remote sensing image [J]. Remote Sensing, 2021, 13(2). DOI: 10.3390/rs13020294.
- [10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 2017: 5999-6009.
- [11] Liu J, Huang H, Sun H, et al. LRAD-Net: an improved lightweight network for building extraction from remote sensing images [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 16: 675-687.
- [12] Cai J, Chen Y. MHA-Net: multipath hybrid attention network for building footprint extraction from high-resolution remote sensing imagery [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5807-5817.
- [13] 曲海成,梁旭. 融合混合注意力机制与多尺度特征增强的高分影像建筑物提取 [J]. 自然资源遥感, 2024, 36(4): 107-117.
Qu Haicheng, Liang Xu. Fusion of hybrid attention mechanism and multi-scale feature enhancement for high-resolution satellite image building extraction [J]. Remote Sensing for Natural Resources, 2024, 36(4): 107-117.
- [14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles: IEEE Computer Society, 2018: 7132-7141.
- [15] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module [J]. Lecture Notes in Computer Science, 2018, 11211: 3-19.
- [16] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles: IEEE Computer Society, 2021: 13713-13722.
- [17] Ding X, Guo Y, Ding G, et al. ACNET: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks [C] // Proceedings of the IEEE International Conference on Computer Vision. Los Angeles: IEEE Computer Society, 2019: 1911-1920.
- [18] 吕少云,李佳田,阿晓荟,等. Res_ASPP_UNet ++: 结合分离卷积与空洞金字塔的遥感影像建筑物提取网络 [J]. 测绘学报, 2023, 27(2): 502-519.
Lü Shaoyun, Li Jiatian, A Xiaohui, et al. Res_ASPP_UNet ++: building an extraction network from remote sensing imagery combining depthwise separable convolution with atrous spatial pyramid pooling [J]. National Remote Sensing Bulletin, 2023, 27(2): 502-519.
- [19] Huang H, Lin L, Tong R, et al. Unet 3 +: A full-scale connected unet for medical image segmentation [C] // IEEE International

- Conference on Acoustics, Speech, and Signal Processing. Los Angeles: IEEE Computer Society, 2020: 1055-1059.
- [20] De Boer P T, Kroese D P, Mannor S, et al. A tutorial on the cross-entropy method[J]. *Annals of Operations Research*, 2005, 134: 19-67.
- [21] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers. Los Angeles: IEEE Computer Society, 2003: 1398-1402.
- [22] Mattyus G, Luo W, Urtasun R. Deeproadmapper: extracting road topology from aerial images[C]// Proceedings of the IEEE International Conference on Computer Vision. Los Angeles: IEEE Computer Society, 2017: 3458-3466.
- [23] 吴开顺, 郑道远, 陈妍伶, 等. 中国典型城市建筑物实例数据集[J]. *中国科学数据*, 2021, 6(1): 182-190.
Wu Kaishun, Zheng Daoyuan, Chen Yanling, et al. A dataset of building instances of typical cities in China [J]. *Science Data Bank*, 2021, 6(1): 182-190.