引 用 格 式：HU Tengyun, XIE Pengfei, WEN Yanan, *et al*. Research on building footprints extraction methods based on different deep learning models[J].Remote Sensing Technology and Application,2023,38(4):892-902.[胡腾云,解鹏飞,温亚楠,等.基于不同深度学习模型提取建筑轮廓的方法研究[J].遥感技术与应用,2023,38(4):892-902.]

DOI:10.11873/j.issn.1004-0323.2023.4.0892

基于不同深度学习模型提取建筑轮廓的方法研究

胡腾云[1],解鹏飞[1,2],温亚楠[3],慕号伟[3]

(1.北京市城市规划设计研究院,北京 100045;
2.北京城垣数字科技有限责任公司,北京 100045;
3.中国农业大学土地科学与技术学院,北京 100083)

摘要：建筑是城市精细化管理的基础单元,利用高分遥感影像快速准确地提取城市建筑轮廓信息对于城市规划及管理有着重要意义。研究基于北京二号高分辨率(0.8 m)遥感数据,建立了北京市建筑轮廓样本库,利用多种语义分割模型U-Net、DANet、UA-Net(U Attention Net)和实例分割模型Mask R-CNN、Mask R-CNN FPN、Mask R-CNN RX FPN来提取城市建筑轮廓并开展精度评价,通过对比不同类型建筑(如楼房、别墅及村庄建筑等)的提取效果,最终选择整体精度最高且提取效果最好的U-Net模型提取了北京市域的所有建筑轮廓。结果表明:U-Net、DANet、UA-Net、Mask R-CNN、Mask R-CNN FPN和Mask R-CNN RX FPN模型的分类精度分别为79.37%、65.59%、71.03%、61.82%、52.53%和59.70%,且U-Net模型训练时间相对较少。U-Net模型对于建筑轮廓的提取有良好的表现;对比不同模型的识别效果发现,语义分割模型对于平房型建筑识别较有优势,实例分割模型则适用于提取城区及周边地区独栋楼房别墅的建筑轮廓,这为开展典型建筑轮廓提取任务的模型选择提供了科学依据,并且识别的城市建筑成果在一定程度上解决了城市内部精细尺度研究数据缺失的问题。

关 键 词：高分辨率遥感影像;建筑轮廓提取;深度学习;语义分割;实例分割

中图分类号：P237;TP391.41 **文献标志码**：A **文章编号**：1004-0323(2023)04-0892-11

1 引 言

第七次全国人口普查显示我国城镇化率目前已超过60%[1]。相应地,我国大城市陆续进入存量甚至减量发展阶段,城市发展的首要任务也从空间增长和生产效率提升转换到城市有机更新与精细化管理[2]。城市建筑作为城市更新的基本空间单元,基于此所获取的建筑密度、建筑高度、容积率等指标也是城市有机更新与精细化管理研究的重要特征,如老城区的建筑高度及城市天际线控制引导等研究工作[3]以及建筑轮廓数据在城市精细化管理、城市重点地区设计中的重要作用[4-5]。当前建筑轮廓数据主要来源于人机交互解译或高精度的测绘数据,其耗时耗力、成本高及更新频率低等问题对开展深入广泛的研究造成了困扰[6]。因此,从遥感影像中快速、准确、实时地获取建筑物轮廓信息,对城市规划及精细化管理有着重要意义。

基于遥感影像提取建筑物轮廓的相关研究始于20世纪80年代,提取方式从早期的人工目视解译、半自动的机器学习[7-11],逐步发展到基于深度学习的全自动提取流程[12]。特别是卷积神经网络

收稿日期:2022-03-21;**修订日期**:2023-07-11

基金项目:北京科技计划"首都城市安全综合风险评估的关键技术研发和示范"(Z211100004121014)。

作者简介:胡腾云(1991—),女,山东烟台人,高级工程师,主要从事城市规划、城市遥感研究。E-mail:hutengyun88@163.com

通讯作者:解鹏飞(1994—),男,河北石家庄人,助理工程师,主要从事遥感信息提取与应用研究。E-mail:x506106211@126.com

(Convolutional Neural Networks,CNNs)的发展,推动了计算机视觉和遥感图像处理领域的范式转变[13]。通过不同的卷积核大小(接受域)学习丰富的上下文信息,以CNNs为基础的发展模型可以在不需要任何先验知识的情况下从输入数据中自动学习层次语义相关表示[14]。

当前用于建筑物轮廓提取的深度学习方法主要有语义分割模型和实例分割模型两大类。语义分割模型是为图像中的每个像素分配标签的过程,使具有相同标签的像素就某种视觉或语义属性相互连接[15]。作为分割精度和分割效率两个方面都表现良好的算法,语义分割模型被广泛应用于遥感影像提取建筑物轮廓中。很多学者通过雷达数据和高分影像等数据源的融合[16-17],深度学习模型的接受域和池化操作调整[18]、低层特征融合[19]、不同网络结构及注意力机制和高斯金字塔的修改等网络结构[20-23]的改进,增加融合策略[24]、全连接条件随机场[25-26]等分割结果的后处理,在不同程度上提升了建筑物轮廓提取的效率和表现。此外,语义分割网络结合二值与浮点数混用可以解决全局网络精度差、训练慢的问题,为建筑物轮廓的实时检测提供了条件[27]。

实例分割模型是将场景图像分割成多个区域(或对象),且注明该区域的类别标签[28]。相较于语义分割模型而言,实例分割模型更接近于人类对世界的认知,而且允许对场景构成元素直接进行后续处理。有学者以三重标记掩码和Xception模块对U-Net语义分割模型进行改进[29-30],实现了建筑轮廓的实例分割任务。典型的实例分割模型Mask R-CNN框架提出之后,学者们利用多尺度空间-光谱特征融合[31]、跳连融合图像特征[32]及增加卷积和mask分支[33]等对其进行调整,提升了建筑轮廓提取的精度。

当前,通过高分遥感影像开展建筑轮廓提取的深度学习网络主要为单一改进语义分割模型或实例分割模型。但是,鲜有研究对比两类深度学习模型在建筑物轮廓提取的表现性及敏感性;两类模型的精度比较也需要从建筑功能、建筑布局等角度开展更为详细的研究。针对上述问题,本研究以北京市全域建筑空间为研究对象,选用经典的语义分割模型U-Net、DANet及其改进模型UA-Net(U Attention Net)和不同的骨干网络的实例分割模型Mask R-CNN、Mask R-CNN FPN及Mask R-CNN RX FPN分别进行高分影像的建筑轮廓提取测试,对语义分割及实例分割模型在建筑轮廓提取任务的精度及效果进行对比,分析两种模型对于不同类型的建筑识别效果的优劣性,并选择最优模型对北京市域建筑空间进行了提取。

2　研究区与数据

作为我国的首都和直辖市,北京市行政区面积达16 410 km^2,位于华北平原北部(115.7°～117.4° E,39.4°～41.6° N),与天津市接壤,环绕于河北省之中。

本研究以2018年为基准年,建立北京市典型区域建筑类样本库。训练样本是多种监督学习(如机器学习、深度学习)算法的重要数据基础,训练样本的质量与数量会对分类精度产生直接影响。本研究选用北京二号高分辨率影像数据,通过数据融合的方式,得到包含3个波段信息(红、绿、蓝波段)的高分辨率影像(0.8 m)。为便于样本库的建立和模型全域推广,研究对高分辨率遥感数据进行1 km×1 km切片处理,并从中随机挑选839张样本影像(约为全域面积的5%)进行建筑轮廓标记。这些建筑样本既有位于山区也有分布于平原区的建成区,包括了高层及低层建筑(图1),保证了样本随机性和多样性。研究通过人机交互的方式对样本的建筑轮廓进行解译,构建北京市高分影像建筑样本库。通过旋转、镜像等图像几何变换操作对样本数据进行扩充,为后续多种深度学习模型的比较提供丰富的训练基础。

3　研究方法

研究基于深度学习提取高分遥感影像建筑轮廓的方法对比的技术路线如图2所示。首先,基于高分遥感影像进行北京市的建筑轮廓样本制备;其次,不同深度学习模型训练调优,利用典型深度学习语义分割,如编码解码结构的U-Net、自注意力机制的DANet、结合自注意力机制和编码解码结构的UA-Net,以及实例分割模型如Mask R-CNN、增加信息特征金字塔Mask R-CNN FPN、替换骨干网络的Mask R-CNN RX FPN等两个类别的多种模型开展迭代训练调优;最后,从多个角度对以上模型提取建筑的精度进行对比评价,并选取提取结果表现最好的模型开展北京市的建筑轮廓识别。

3.1　语义分割模型

语义分割旨在实现图像像素级分类,即为图像

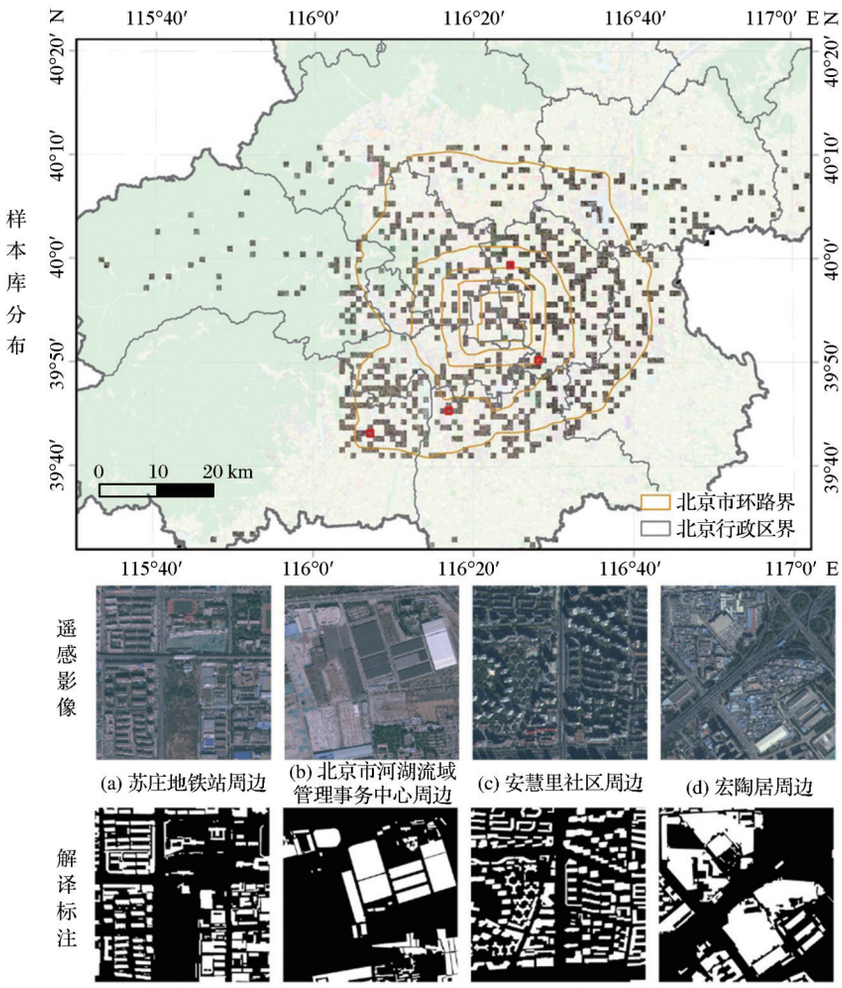


图 1 样本库分布及遥感影像与解译样本(包括中高层建筑和平房区等)

Fig.1 Sample database distribution and examples of remote sensing imagery and interpretation samples

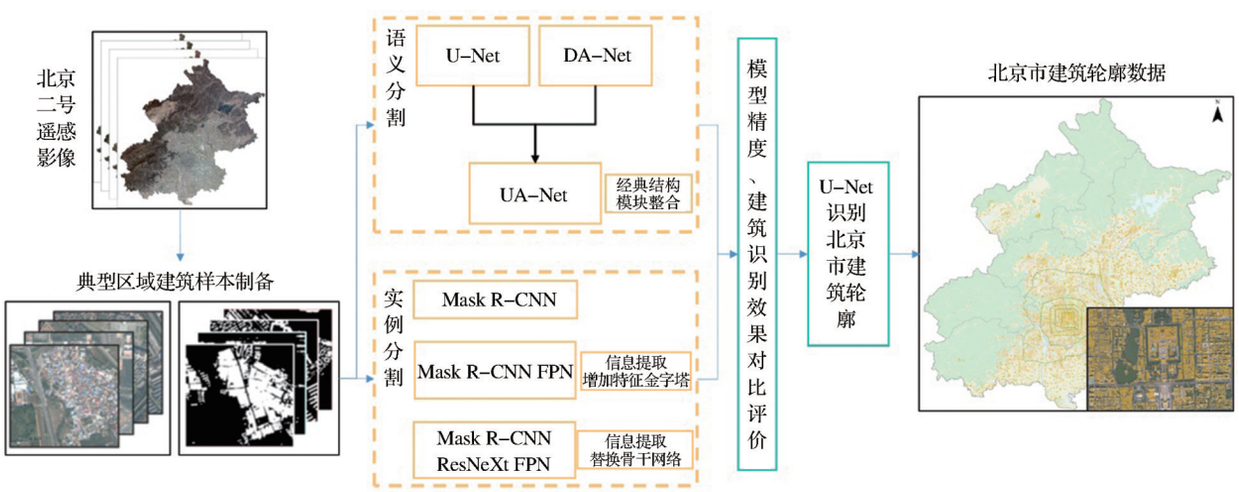


图 2 深度学习提取建筑轮廓方法对比技术路线

Fig.2 The flow chart of Comparison of deep learning methods for extracting building footprints

中的每个像素分配类别标签。U-Net[34]为典型的编码器-解码器结构,其对应的特征拼接可以有效提升模型精度。另外,结构简单且参数轻量的U-Net在设计网络方面有优势。DANet创新性地提出了注意力机制进行语义信息的提取,有效提升了精度[35]。

U-Net网络最早用于生物和医学影像的分割,是一种经典的语义分割算法,近年来在目标分类方面得到了广泛应用。U-Net网络结构主要由卷积层、最大池化层(下采样)、反卷积层(上采样)以及ReLU(Rectified Linear Unit,修正线性单元)非线性激活函数组成,为典型的编码器-解码器U型结构。U-Net网络具有利用少量样本数据训练学习的能力以及轻量化的网络模型方便对其进行改进。

Dual Attention Network(DANet)利用自注意力机制进行丰富语义信息的捕获,在带有空洞卷积的FCN架构的尾部添加两个并行的注意力机制:空间注意力机制和通道注意力机制。并且注意力机制模块为独立模块,可以直接插入到已有的FCN中,不会增加太多参数,有效地增强特征表示。

基于U-Net编码器—解码器及DANet注意力机制的特点,本研究设计了U-Attention Net(UA-Net)网络结构(如图3)。该网络主要分为3个部分,左侧为5层卷积层,右侧对应5层上采样,在两者连接中添加了注意力机制。左侧的卷积层进行提取和学习图像特征,右侧的上采样层与左侧对应,用于还原至原图像大小,在上采样的同时会连接对应卷积层的特征图,得到更高层次的特征组合,从而提高模型的结果精确度。注意力机制中包含了空间注意力和通道注意力机制。空间注意力机制是任意空间位置特征的更新,即通过图像所有空间位置上特征的加权聚合进行更新,权重由两个空间位置上的特征相似性决定,特征相似则权重更高,不需要考虑两个空间位置的距离;通道注意力机制中应用了相似的自注意力机制来学习任意两个通道映射之间的关系,通过所有通道的加权聚合来更新某个通道。

3.2　实例分割模型

Mask R-CNN[36]是一种灵活的实例分割框架,可以对骨干网络进行修改替换以及增加不同的分支完成不同的任务,包括了骨干网络、ROI Align层和全卷积网络层。

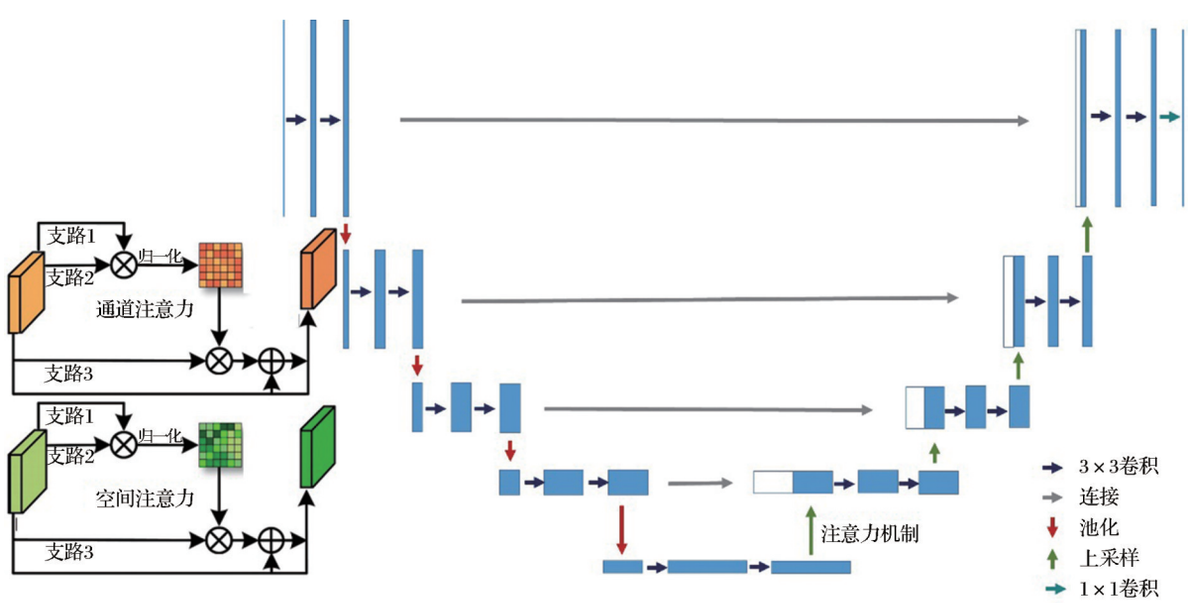


图3　UA-Net网络结构图

Fig.3　Structure of UA-Net

初始的Mask R-CNN模型以深度残差网络[37](ResNet)为骨干网络,骨干网络ResNeXt结构块[38]借鉴了ResNet中的恒等映射与残差学习的思路,将残差学习部分单路卷积变成了多个支路的多路卷积,多路径分组卷积遵循分割、转换合并范式,各个分支同构使用相同的拓扑结构。ResNeXt作为最强骨干网络之一,可以在不增加参数复杂度的前提下提高准确率。考虑到遥感影像中建筑物轮廓的特征复杂,图像金字塔的思想可以解决物体监测场景中小尺寸物体监测困难的问题,因此研究将特征金字塔网络(Feature Pyramid Networks,FPN)[39]引入到模型中,FPN采用自底向上、自顶向下以及横向连接的方式将浅、中层次特征图等不同等级特征图结合在一起,可在提升模型精度的同时,对其训练的时间复杂度增加很少。

Mask R-CNN由Faster R-CNN[40]拓展而来,

在每个ROI(Region of Interest)增加一个预测分割掩码的全卷积网络层,进而在生成检测框的同时,实现目标掩膜预测。ROI Pooling是从每个ROI提取特征图的标准操作,采用取整的方法将ROI量化到和特征图相匹配的粒度。量化操作对像素级的语义分割过程有着较大影响,改进之后提出的ROI Align层,采用双线性插值的方法获取每个位置的精确值,消除了粗糙量化操作带来的误差,同时保证提取特征与像素精准对齐。

全卷积网络产生对应的掩膜预测分支。其中分类预测分支对ROI给出预测,产生对应的矩形框的输出边界;基于输出边界,掩膜预测分支产生二值掩码。Mask R-CNN对于每个分类输出边界均独立地对应掩膜预结果,有效的避免了类间的竞争,提升了模型的性能。本研究选取经典实例分割模型Mask R-CNN(以ResNet为骨干网络),并在其网络结构基础上,分别以残差网络ResNet、ResNeXt结合FPN作为骨干网络衍生出另外两个模型Mask R-CNN FPN(以ResNet和FPN为骨干网络)和Mask R-CNN RX FPN(以ResNeXt和FPN为骨干网络),如图4,测试以上3个实例分割模型对于高分影像提取建筑轮廓的精度与效果。

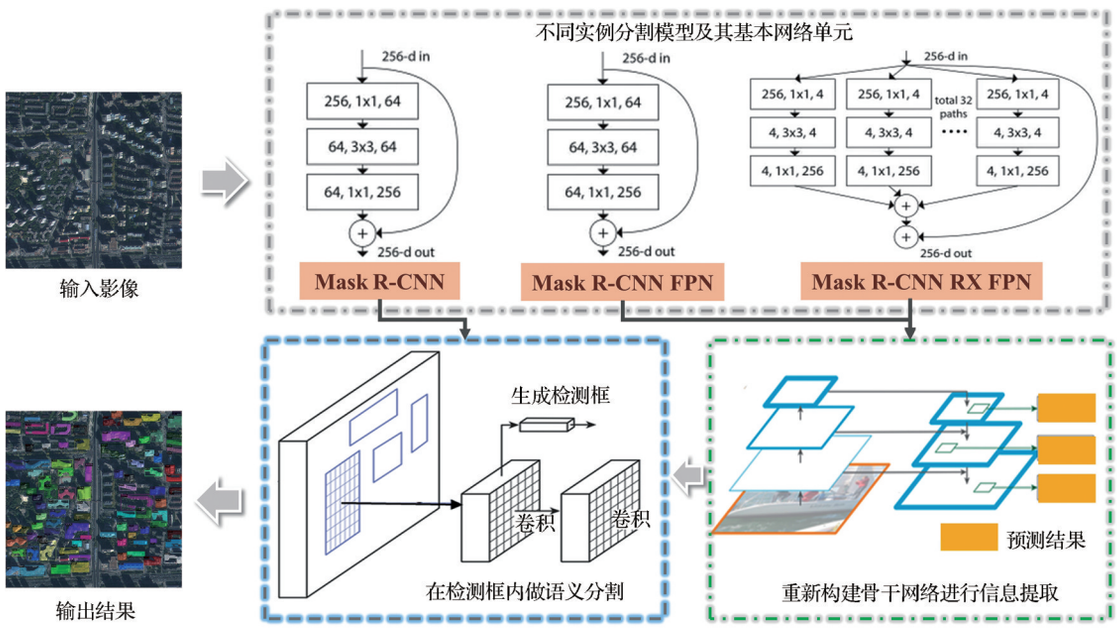


图4 实例分割模型结构图

Fig.4 Structure of instance segmentation model

3.3 模型评价方法

本研究使用4个指标评价模型,分别为精确率(precision)、召回率(recall)、F1指数(F1)以及重叠度(Intersection over Union,IoU),从多个角度对建筑轮廓划分效果进行定量评估。其中,精确率、召回率、F1指数的定义如下:

$$Precision=\frac{TP}{TP+FP} \tag{1}$$

$$Recall=\frac{TP}{TP+FN} \tag{2}$$

$$F1=\frac{2\cdot Precision\cdot Recall}{Precision+Recall} \tag{3}$$

其中:TP表示实际为正例且被模型划分为正例的像素数;FP表示实际为负例但被模型划分为正例的像素数;FN表示实际为正例但被模型划分为负例的像素数。在本研究中,建筑物轮廓的像素为正例,背景像素为负例。

重叠度IoU的定义为:

$$IoU=\frac{|P_p\cap P_t|}{|P_p\cup P_t|} \tag{4}$$

其中,P_p表示预测为建筑物轮廓的像素集,P_t表示样本真实为建筑轮廓的像素集,$|\cdot|$代表计算集合中像素数的函数。

3.4 实验环境及模型测试

为保证实验对比的有效性,研究在统一的硬件及软件环境下选择典型的3种语义分割和3种实例分割模型进行方法对比。软件环境使用开源的PyTorch学习框架,使用Python语言编程实现语义分割、实例分割算法网络,硬件环境为惠普服务器,配

备有 TU102［TITAN RTX］显卡(24 G)、64 位 Ubuntu 18.04操作系统。随机选取样本库中的789张切片(占比95%)作为训练集,其余样本50张为验证集。6种深度学习网络分别为语义分割网络U-Net、DANet、UA-Net,实例分割网络 Mask R-CNN、Mask R-CNN FPN 以及 Mask R-CNN RX FPN进行建筑提取方法对比研究。

4　结果与分析

4.1　多模型精度评价

利用模型评价参数对2类(6种)深度学习模型的验证结果开展模型的精度评价。由表2可知,在当前相同的工作环境(样本量、高分影像精度等),语义分割模型表现了更高的精度。语义分割模型U-Net、DANet 以及 UA-Net 的精确率分别为88.98%、81.68%、83.82%,召回率分别为87.84%、76.58%、82.01%,F1指数分别为88.09%、78.46%、82.30%,IoU 分别为 79.37%、65.59%、71.03%;实例分割模型 Mask R-CNN、Mask R-CNN FPN、Mask R-CNN RX FPN的精确率分别为81.13%、82.71%、82.05%,召回率分别为66.73%、59.02%、64.36%,F1指数分别为72.65%、67.97%、71.35%,IoU分别为57.80%、52.53%、56.44%。总体来看,深度学习模型的整体精度随着网络复杂性增加而降低,其训练时长随着网络的复杂呈现增加趋势。也就是说,U-Net(语义分割)和Mask R-CNN(实例分割)网络本身在各自对应类型内的精度和效率表现上相对较优。

4.2　多模型提取建筑轮廓效果对比

总体来看,相比于实例分割模型,语义分割模型对于建筑轮廓的提取效果较好且较为鲁棒(对各类建筑都能进行明确区分)。以语义分割和实例分割的最优模型U-Net和Mask-R-CNN为例(表2),语义分割获取的建筑物边缘比较平滑且结构清晰,而实例分割则有明显的边缘不清晰问题(图5(a)、图5(b);图6(b)、图6(c))。然而,语义分割将部分高架桥区域错分为建筑,实例分割的这一问题相对不突出(图5(c)、图5(d))。

表1　不同模型比较的参数设置

Table 1　Parameter settings for different models

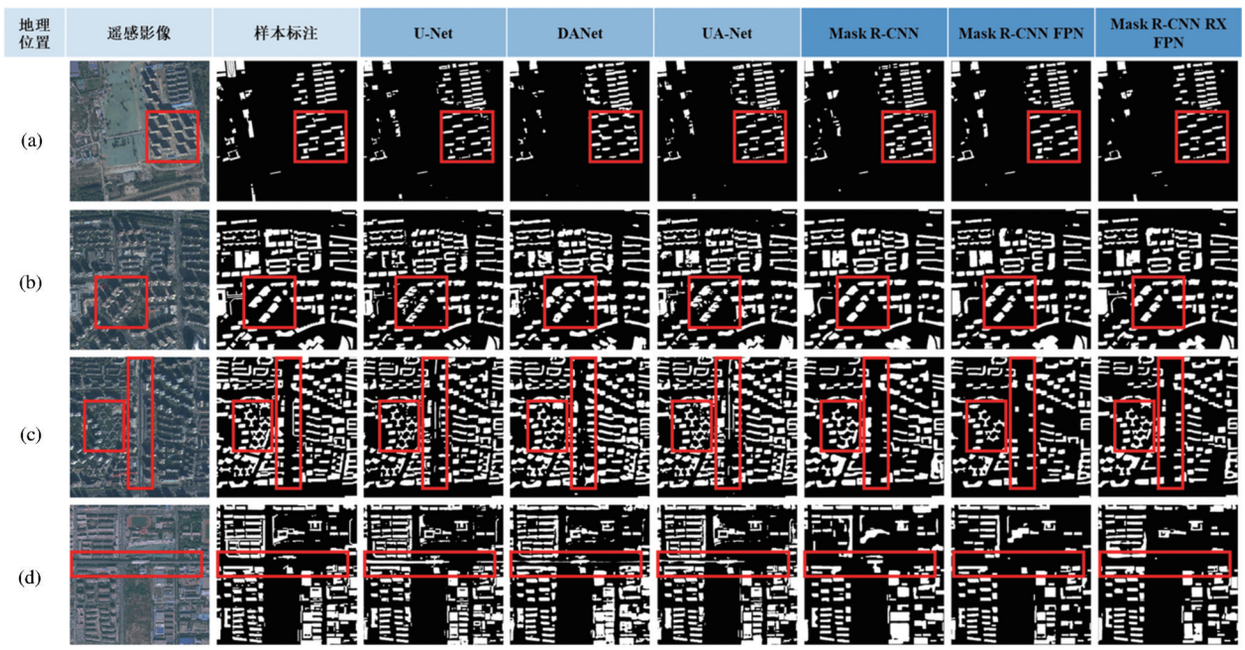
语义分割模型		实例分割模型	
训练参数	参数值	训练参数	参数值
初始学习率	0.001	初始学习率	0.001
优化器	Adam	优化器	Adam
损失函数	binary cross entropy	损失函数	分类损失、检测损失和掩膜分割损失之和
批尺寸	10	批尺寸	10
时期	20	时期	20

表2　不同深度学习模型精度对比

Table 2　Accuracy comparison of different deep learning models

	骨干网络(backbone)	精确率(Precision)	召回率(Recall)	F1指数	IoU	训练时长/h
U-Net	--	88.98%	87.84%	88.09%	79.37%	20
DANet	--	81.68%	76.58%	78.46%	65.59%	30
UA-Net	--	83.82%	82.01%	82.30%	71.03%	35
Mask R-CNN	ResNet	80.62%	72.46%	75.34%	61.82%	17
Mask R-CNN FPN	ResNet FPN	83.13%	66.65%	67.97%	52.53%	31
Mask R-CNN RX FPN	ResNeXt FPN	85.14%	66.36%	73.61%	59.70%	43

为了验证模型的鲁棒性,研究对比了模型在不同建筑类型间的提取效果,如城市内部的高层及中层建筑与城市周边乡村平房区等。因为语义分割模型对具有明显的建筑间距和空间形态特征的建筑敏感性较强,U-Net模型提取的城市内部居住类建筑轮廓较为准确且建筑边缘也较完整(图5(a)、图5(b)),但对于建筑阴影仍存在部分漏分现象,特别是中低层建筑(图5(c))。虽然DANet模型的整体表现略低于U-Net,但由于注意力机制的引入,能够弥补由于影像阴影所造成的高层建筑部分漏分现象,从而提高对模型的表现(图5(b)、图5(c))。类似的模型表现也适用于UA-Net。Mask R-CNN



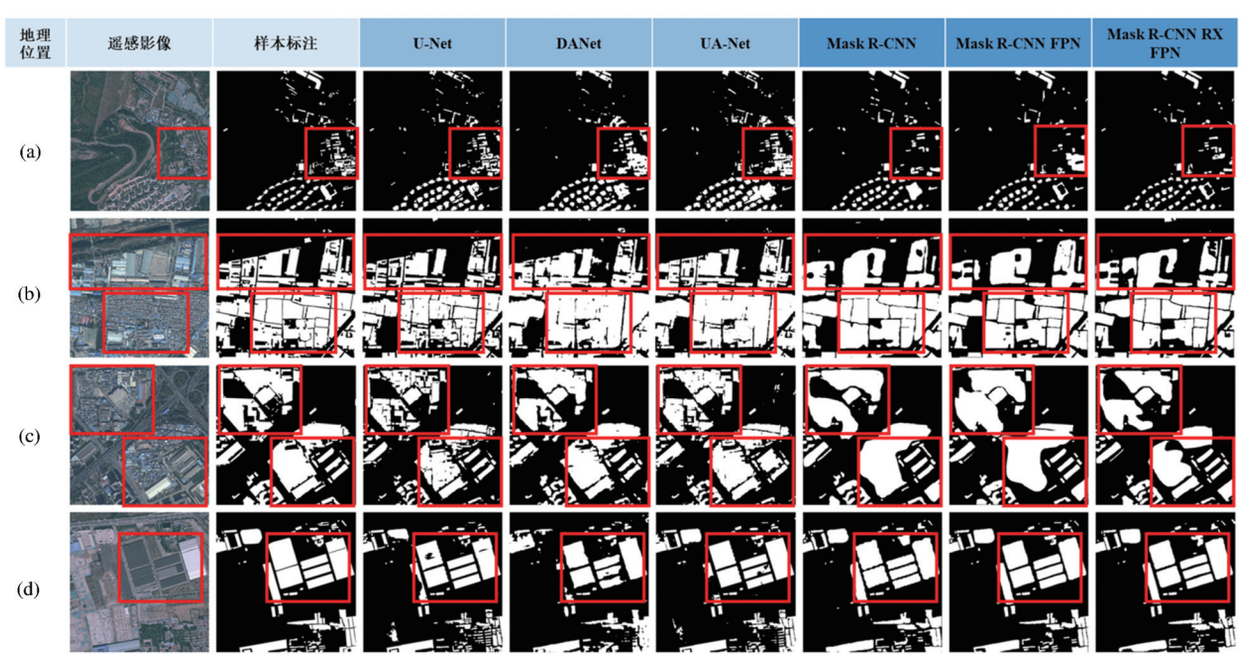
(a)贺村新村(40.058267 N, 116.386069 E);(b)望京西园-三区(40.000311 N,116.478681 E);(c)安慧里社区(39.994333 N,116.413703 E);(d)苏庄地铁站(39.723286 N,116.126764 E)

图 5 不同模型的城市内部建筑物轮廓提取结果对比

Fig.5 Comparison of building footprints extraction in city with different models

模型对于高层建筑群独栋建筑轮廓的提取置信度较高(图 5(a)),但存在提取的建筑轮廓边缘模糊情况(图 5(b)、图 5(c));对于中低层建筑,这种由于边缘模糊所导致的提取效果变差显得尤为明显(图 5(c))。类似的问题也见于 Mask R-CNN FPN 和 Mask R-CNN RX FPN 模型。

对于城市周边的乡村平房,语义分割模型(例如 U-Net)的表现主要受楼房之间的间隙所限,因而



(a)中骏绿洲庄园(39.918202 N,116.451586 E);(b)大一村(39.838892 N,116.595497 E);(c)宏陶居(39.961353 N,116.398486 E);(d)河湖流域管理事务中心(39.907803 N,116.297894 E)

图 6 不同模型的城市周边乡村平房轮廓提取结果对比

Fig.6 Comparison of building footprints extraction in village with different models

会把间隙较小的多个建筑提取为连片的对象。相比于城市内居住类建筑的提取,其模型表现效果相对较差。例如,大片密集的农村居民点建筑在语义分割模型中往往被识别为大的片状建筑(图6(b)、图6(c))。针对乡村地区比较大的平房(或厂房)区域,语义分割和实例分割模型的效果均表现较好,且实例分割相对更优,其对于建筑对象的识别也更为准确(图6(d))。

深度学习模型对城市内、外建筑识别表现之间的差异主要受两个因素影响。一方面,在构建训练样本库阶段是以城市建筑为主,对于农村比较密集且较小的居民点建筑缺乏精细的样本进行区分。另一方面,受影像分辨率影响,农村密集的居民点建筑本身的间隙较小(例如小于1米),在影像上视觉表现模糊,模型无法识别并提取。鉴于以上分析,研究认为基于更高分辨率的遥感数据是解决农村地区密集居民点建筑提取的发展方向。

4.3　北京市域建筑轮廓提取结果评估

基于上述分析,利用提取效果表现较好的U-Net模型对2018年北京市域遥感影像进行建筑轮廓提取,并与相同年份的测绘建筑数据成果进行对比,评估模型在北京市域的适用性及有效性。研究发现U-Net模型对于胡同、高层建筑及别墅等建筑类型均能够进行有效提取,边缘明确,与测绘建筑数据效果较为相近。根据统计,U-Net模型提取北京市建筑投影面积为1 067.3 km^2,测绘建筑数据投影面积为726.4 km^2,二者相交建筑投影面积为615.4 km^2,即测绘建筑数据和深度学习模型识别的建筑轮廓数据均为建筑投影的面积为615.4 km^2,占测绘建筑数据的84%,由此可见,U-Net模型对于测绘建筑数据有较高的召回率。虽然U-Net模型本身得到的结果存在一定程度的分类误差,其他因素也会对二者的面积差异产生一定影响。例如测绘建筑数据的阶段性更新特性使得部分地区建筑更新未能进行有效监测;测绘建筑数据对农村居民点建筑的覆盖性相对较弱;采用的遥感影像非正射投影也会增加了U-Net识别的建筑投影面积等。综上所述,U-Net模型在大范围建筑的识别结果与当前的测绘数据有较高的一致性,并且提取效率较高,为建筑的快速提取和实时监测提供了技术手段。

5　结　论

本研究基于北京2号高分遥感影像,建立了北京市典型建筑轮廓的样本库,并利用多种语义分割、实例分割模型进行建筑轮廓提取的对比,评估不同模型在城市范围内建筑识别提取的优劣。通过对比评估可以看出,在城市建筑轮廓提取中,网络层数较浅的模型在识别遥感影像中的建筑颜色、纹理、边缘等特征有明显优势。整体而言,语义分割模型在本实验中相比于实例分割模型更具优势,特别是模型应用于不同类型建筑的泛化性和鲁棒性。例如,语义分割模型提取的建筑边缘表征上更加清晰。

基于对不同类型深度学习模型的对比,研究选择了在当前样本情况下,精度表现较高的语义分割模型U-Net对北京市建筑轮廓进行提取,形成了一套精度较高的北京市建筑轮廓数据,为大范围城市建筑轮廓的高效快速提取与深度学习模型提取不同类型建筑研究提供了可靠的技术方案。

利用深度学习模型对高分遥感影像进行建筑轮廓提取,可以满足地图制图和地理信息系统对建筑物轮廓数据采集和自动更新的需求。本研究对不同种类、不同模型提取不同类型建筑的精度进行评估,这有利于在不同建筑提取任务中高效地选择更有针对性的模型,有利于遥感影像在城市规划、智慧城市建设等领域的深入应用和扩展,对遥感影像制图、城市变化监测、三维建模、地理信息系统的数据获取、城市空间数据库的更新等建设“数字化城市”领域具有重要的应用价值。

参考文献(References):

[1] National Bureau of Statistics, Office of the Seventh National Census Leading Group of The State Council. Bulletin of the Seventh National Population Census (No. 7) - Urban and Rural Population and Floating Population Situation [EB/OL], http: //www. stats. gov. cn/tjsj/tjgb/ rkpcgb/qgrkpcgb/202106/t20210628_1818826.html, 2021, 2022.[国家统计局,国务院第七次全国人口普查领导小组办公室.第七次全国人口普查公报(第七号)—城乡人口和流动人口情况[EB/OL], http://www.stats.gov.cn/tjsj/tjgb/rkpcgb/qgrkpcgb/202106/t20210628_1818826.html, 2021, 2022.]

[2] WU Jiang. Organic regeneration and elaborated urban management[J]. Time Architecture,2021,180(4):6-11.[伍江.城市有机更新与精细化管理[J].时代建筑,2021,180(4):6-11.]

[3] WANG Qian, WANG Zhenmao, YANG Yifan. Oriented to fine management of urban building height guidance——Urban design practice of building height and skyline in the old town of Huairou New Town in Beijing[J]. Beijing Planning Review,

2017,176(5):71-76.[王倩, 王振茂, 杨一帆. 面向精细化管理的城市建筑高度引导——北京怀柔新城老城区建筑高度及天际线城市设计实践[J]. 北京规划建设,2017,176(5):71-76.]

[4] ZHU He, TANG Yan. Central district urban design for detailed management in new institutional environment: Zhongguancun area, Beijing Case[J]. Planners, 2017, 33(10): 17-23.[祝贺, 唐燕. 新制度环境下对接精细化管理的重点地区城市设计—以北京中关村大街地区城市设计为例[J]. 规划师,2017,33(10):17-23.]

[5] WU Chen, LI Jing, ZHOU Xiaojie, *et al*. Research and practice of urban fine design management——Take Beijing Mentougou New Town and Fangshan New Town urban design pilot work for example[J]. Beijing Planning Review, 2018, 179(2): 186-191.[吴晨, 李婧, 周小洁,等. 城市精细化设计管理研究及实践——以北京门头沟新城及房山新城城市设计试点工作为例[J]. 北京规划建设,2018,179(2):186-191.]

[6] ZHANG Qingyun, ZHAO Dong. Research on methods of building extraction from high resolution remote sensing images [J]. Geomatics & Spatial Information Technology, 2015, 38(4):74-78.[张庆云, 赵冬. 高空间分辨率遥感影像建筑物提取方法综述[J]. 测绘与空间地理信息,2015,38(4):74-78.]

[7] HUANG X, ZHANG L, ZHU T. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2013, 7(1): 105-115. DOI: 10.1109/JSTARS. 2013. 2252423

[8] OK A O, SENARAS C, YUKSEL B. Automated detection of arbitrarily shaped buildings in complex environments from monocular vhr optical satellite imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 51(3): 1701-1717. DOI: 10.1109/TGRS.2012. 2207123

[9] CHEN R, LI X, LI J. Object-based features for house detection from rgb high-resolution images[J]. Remote Sensing, 2018, 10(3): 451. DOI: 10.3390/rs10030451

[10] FANG Xin, CHEN Shanxiong. High resolution remote sensing image building extraction in dense urban areas[J]. Bulletin of surveying and mapping, 2019, 505(4): 79-83.[方鑫, 陈善雄. 密集城区高分辨率遥感影像建筑物提取[J]. 测绘通报, 2019,505(4):79-83.]

[11] HUANG X, YUAN W L, LI J Y, *et al*. A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 10(2): 654-668. DOI: 10.1109/JSTARS.2016. 2587324

[12] PILINJA SUBRAHMANYA P, HARIDAS AITHAL B, MITRA S. Automatic extraction of buildings from UAV-based imagery using Artificial Neural Networks[J]. Journal of the Indian Society of Remote Sensing, 2021, 49: 681-687. DOI: 10.1007/s12524-020-01235-z

[13] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521(7553): 436-444. DOI: 10.1038/nature14539

[14] MARMANIS D, DATCU M, ESCH T, *et al*. Deep learning earth observation classification using imagenet pretrained networks[J]. IEEE Geoscience & Remote Sensing Letters, 2015, 13(1): 105-109. DOI: 10.1109/LGRS.2015.2499239

[15] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, *et al*. Simultaneous detection and segmentation[C]//Computer Vision - ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13. Springer International Publishing, 2014: 297-312. DOI: 10.1007/978-3-319-10584-0_20

[16] GUO Feng, MAO Zhengyuan, ZOU Weibin, *et al*. A method for building extraction by fusing feature information from LiDAR data and high-resolution imagery[J]. Journal of Geo-information Science, 2020, 22(8): 1654-1665.[郭峰, 毛政元, 邹为彬,等. 融合LiDAR数据与高分影像特征信息的建筑物提取方法[J]. 地球信息科学学报,2020,22(8):1654-1665.]

[17] YANG Jianyu, ZHOU Zhenxu, DU Zhenrong, *et al*. Rural construction land extraction from high spatial resolution remote sensing image based on SegNet semantic segmentation model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(5): 251-258.[杨建宇, 周振旭, 杜贞容,等. 基于SegNet语义模型的高分辨率遥感影像农村建设用地提取[J]. 农业工程学报, 2019,35(5):251-258.]

[18] YANG Xubo, TIAN Jinwen. Research on small building extraction method in small dataset[J]. Bulletin of surveying and mapping, 2019, 511(10): 51-55.[杨旭勃, 田金文. 小数据集中的小型建筑物提取方法研究[J]. 测绘通报, 2019, 511(10):51-55.]

[19] REN Xinlei, WANG Yangping, YANG Jingyu, *et al*. Building detection from remote sensing images based on improved U-net[J]. Laser & Optoelectronics Progress, 2019, 56(22): 195-202.[任欣磊, 王阳萍, 杨景玉,等. 基于改进U-net的遥感影像建筑物提取[J]. 激光与光电子学进展,2019,56(22): 195-202.]

[20] GUO M, LIU H, XU Y, *et al*. Building extraction based on Unet with an attention block and multiple losses[J]. Remote Sensing, 2020, 12(9): 1400. DOI: 10.3390/rs12091400

[21] XU Zhaohong, LIU Yu, QUAN Jicheng, *et al*. Buildings egmentation of remote sensing images based on VGG16 pre-encoding [J]. Science Technology and Engineering, 2019, 19(17): 250-255.[徐昭洪, 刘宇, 全吉成,等. 基于VGG16预编码的遥感图像建筑物语义分割[J]. 科学技术与工程,

2019,19(17):250-255.]

[22] XIE Yuehui, LI Baishou, LIU Congna. Urban building extraction by combining multiple image features and CNN[J]. Remote Sensing Information,2020,35(5):80-88.[谢跃辉, 李百寿, 刘聪娜. 结合多种影像特征与CNN的城市建筑物提取[J]. 遥感信息,2020,35(5):80-88.]

[23] LIU Wenxiang, SHU Yuanzhong, TANG Xiaomin, *et al*. 2020. Remote sensing image segmentation using dual attention mechanism Deeplabv3+algorithm[J]. Tropical Geography, 2020,40(2):303-313.[刘文祥, 舒远仲, 唐小敏,等. 采用双注意力机制Deeplabv3+算法的遥感影像语义分割[J]. 热带地理,2020,40(2):303-313.]

[24] LIU X. The application of deep convolution neural network to building extraction in remote sensing images[J]. World Scientific Research Journal, 2020, 6(3): 136-144. DOI: 10.6911/WSRJ.202003_6(3).0017

[25] ZHANG Haoran, ZHAO Jianghong, ZHANG Xiaoguang. High-resolution image building extraction using U-net Neural Network[J]. Remote Sensing Information,2020,35(3):143-150.[张浩然, 赵江洪, 张晓光. 利用U-net网络的高分遥感影像建筑提取方法[J]. 遥感信息,2020,35(3):143-150.]

[26] WANG Junqiang, LI Jiansheng, ZHOU Huachun, *et al*. Typical element extraction method of remote sensing image based on Deeplabv3 + and CRF[J]. Computer Engineering,2019, 45(10):260-265,271.[王俊强, 李建胜, 周华春,等. 基于Deeplabv3+与CRF的遥感影像典型要素提取方法[J]. 计算机工程,2019,45(10):260-265,271.]

[27] ZHU Tianyou, DONG Feng, GONG Huixing. Remote sensing building detection based on binarized semantic segmentation[J]. Acta Optica Sinica,2019,39(12):372-382.[朱天佑, 董峰, 龚惠兴. 基于二值语义分割网络的遥感建筑物检测[J]. 光学学报,2019,39(12):372-382.]

[28] SILBERMAN N, SONTAG D, FERGUS R. Instance segmentation of indoor scenes using a coverage loss[C]//Computer Vision - ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13. Springer International Publishing, 2014: 616-631. DOI: 10.1007/978-3-319-10590-1_40

[29] WAGNER F H, DALAGNOL R, TARABALKA Y, *et al*. U-Net-Id, an instance segmentation model for building extraction from satellite images—Case study in the Joanópolis City, Brazil[J]. Remote Sensing, 2020, 12(10): 1544. DOI: 10.3390/rs12101544

[30] HUI Jian, QIN Qiming, XU Wei, *et al*. Instance segmentation of buildings from high-resolution remote sensing images with Multitask Learning[J]. Acta Scientiarum Naturalium Universitatis Pekinensis,2019,55(6):1067-1077.[惠健, 秦其明, 许伟等. 基于多任务学习的高分辨率遥感影像建筑实例分割[J]. 北京大学学报(自然科学版),2019,55(6):1067-1077.]

[31] SONG Shiran. Research on object recognition method of urban buildings in high spatial resolution remote sensing imagery [D]. Beijing: Beijing University of Civil Engineering and Architecture,2020.[宋师然. 高分遥感城市建筑物对象化识别方法研究[D]. 北京:北京建筑大学,2020.]

[32] LI Sensen, WU Qing. Multi-target detection and segmentation of remote sensing images based on improved Mask RCNN[J]. Computer Engineering and Applications, 2020, 56(14): 183-190.[李森森, 吴清. 改进Mask R-CNN的遥感图像多目标检测与分割[J]. 计算机工程与应用,2020,56(14):183-190.]

[33] HU Minjun, FENG Dejun, LI Qiang. Automatic extraction of buildings based on instance segmentation model[J]. Bulletin of Surveying and Mapping,2020,517(4):16-20,62.[瑚敏君, 冯德俊, 李强. 基于实例分割模型的建筑物自动提取[J]. 测绘通报,2020,517(4):16-20,62.]

[34] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241. DOI: 10.1007/978-3-319-24574-4_28

[35] FU J, LIU J, TIAN H, *et al*. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.

[36] HE K, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017:2961-2969.

[37] HE K, ZHANG X, REN S, *et al*. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.

[38] XIE S, GIRSHICK R, DOLLÁR P, *et al*. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:1492-1500.

[39] LIN T Y, DOLLÁR P, GIRSHICK R, *et al*. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:2117-2125.

[40] REN S, HE K, GIRSHICK R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28.

Research on Building Footprints Extraction Methods based on Different Deep Learning Models

HU Tengyun[1], XIE Pengfei[1,2], WEN Yanan[3], MU Haowei[3]

(1.*Beijing Municipal Institute of City Planning and Design*, *Beijing* 100045, *China*;
2.*Beijing City Interface Technology Limited Liability Company*, *Beijing* 100045, *China*;
3.*College of Land Science and Technology*, *China Agricultural University*, *Beijing* 10083, *China*)

Abstract: Building is the basic unit of urban refined management, the rapid and accurate extraction of urban building footprints based on high-resolution remote sensing images is of great significance for urban planning and management. Based on the high-resolution (0.8 m) remote sensing data of Beijing-2, a sample library of building footprints in Beijing was established. We used multiple semantic segmentation models, U-Net, DANet, UA-Net (U Attention Net) and instance segmentation models, Mask R-CNN, Mask R-CNN FPN, Mask R-CNN RX FPN to extract building footprints, performed accuracy evaluation and compare the extraction effects of different types of buildings (such as buildings, villas and village buildings, etc.). Finally, we selected the U-Net model with the highest overall accuracy and the best extraction performance to extract all building footprints in the Beijing area. The results show that the classification accuracy of U-Net, DANet, UA-Net, Mask R-CNN, Mask R-CNN FPN and Mask R-CNN RX FPN models are 79.37%, 65.59%, 71.03%, 61.82%, 52.53% and 59.70%, respectively. And the U-Net model training time is relatively short. The U-Net has a good performance for the extraction of building footprints. Comparing the recognition effects of different models, it is found that the semantic segmentation model is more advantageous for the recognition of bungalow buildings, while the instance segmentation model is suitable for single-family buildings and villas in urban and surrounding areas. The study provides a scientific basis for model selection for typical building footprints extraction tasks and our achievement solves the problem of lack of fine-scale research data in cities to a certain extent.

Key words: High-resolution remote sensing images; Building footprints extraction; Deep learning; Semantic segmentation; Instance segmentation