

基于改进 Mask-RCNN 的遥感影像建筑物提取^①



何代毅^{1,2,3}, 施文灶^{1,2,3}, 林志斌^{1,2,3}, 乔星星^{1,2,3}, 刘芫汐^{1,2,3}, 林耀辉^{1,2,3}

¹(福建师范大学 福建省光电传感应用工程技术研究中心, 福州 350117)

²(福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)

³(福建师范大学 福建省光子技术重点实验室, 福州 350007)

通讯作者: 施文灶, E-mail: swz@fjnu.edu.cn

摘 要: 由于遥感影像中建筑物种类繁多且与周围环境信息混淆, 传统方法难以实现建筑物的准确高效提取. 本文提出了一种基于改进 Mask-RCNN 的建筑物自动提取方法, 利用 PyTorch 深度学习框架搭建改进 Mask-RCNN 网络模型架构, 在网络的设计中添加了路径聚合网络和特征增强功能, 通过监督和迁移学习的方式对 Inria 航空影像标签数据集进行多线程迭代训练与模型优化学习, 实现了建筑物的自动精确分割和提取. 基于不同开源数据集, 分别与 SVM、FCN、U-net 和 Mask-RCNN 等建筑物提取算法进行对比, 实验表明, 本文方法可以高效准确、高效地提取建筑物, 对于同一个数据集, 提取结果的 *mAP*、*mRecall*、*mPrecision* 和 *F1* 分数这 4 个评价指标均优于对比算法.

关键词: 建筑物提取; Mask-RCNN; PyTorch; 实例分割

引用格式: 何代毅, 施文灶, 林志斌, 乔星星, 刘芫汐, 林耀辉. 基于改进 Mask-RCNN 的遥感影像建筑物提取. 计算机系统应用, 2020, 29(9): 156–163.
<http://www.c-s-a.org.cn/1003-3254/7484.html>

Building Extraction from Remote Sensing Image Based on Improved Mask-RCNN

HE Dai-Yi^{1,2,3}, SHI Wen-Zao^{1,2,3}, LIN Zhi-Bin^{1,2,3}, QIAO Xing-Xing^{1,2,3}, LIU Yuan-Xi^{1,2,3}, LIN Yao-Hui^{1,2,3}

¹(Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fujian Normal University, Fuzhou 350117, China)

²(Key Laboratory of Optoelectronic Science and Technology for Medicine (Ministry of Education), Fujian Normal University, Fuzhou 350007, China)

³(Fujian Provincial Key Laboratory of Photonics Technology, Fuzhou 350007, China)

Abstract: In view of the variety of buildings, and the confusion with the surrounding environment in the remote sensing image, the traditional methods are difficult to extract the buildings efficiently and accurately. This study proposes a method of building automatic extraction method based on the improved Mask-RCNN. In the proposed method, an improved Mask-RCNN network model framework is constructed by using PyTorch deep learning framework, path aggregation network and feature enhancement function are added in the network design, and multi-threaded iterative training and model optimization learning are conducted for the Inria aerial image tag data set by means of supervision and migration learning, thus automatic and accurate segmentation and extraction of buildings are achieved. The proposed method is compared with SVM, FCN, U-net, Mask-RCNN, and other building extraction algorithms with different open-source datasets. The experimental results show that the proposed method has competitive performance, which can extract buildings more efficiently, accurately, and quickly in different open-source datasets, and the four evaluation indexes of *mAP*, *mRecall*, *mPrecision*, and *F1* scores extracted in the same dataset are better than the compared algorithms.

Key words: building extraction; Mask-RCNN; PyTorch; instance segmentation

① 基金项目: 国家自然科学基金青年基金 (41701491); 福建省自然科学基金面上项目 (2017J01464, 2018J01619)

Foundation item: Young Scientists Fund of National Natural Science Foundation of China (41701491); General Program of Natural Science Foundation of Fujian Province, China (2017J01464, 2018J01619)

收稿时间: 2019-12-04; 修改时间: 2019-12-27; 采用时间: 2020-01-07; csa 在线出版时间: 2020-09-04

1 引言

基于遥感影像建筑物提取的基础理论研究始于20世纪80年代,随着遥感卫星技术的进步,高分辨率遥感影像在地面目标地自动提取中得到了广泛应用.高分辨率遥感影像具有更加精细丰富的信息,建筑物是其中最普遍且最复杂的地物信息之一.建筑物提取是城市遥感影像的主要研究方向,一方面,建筑物的精确提取有利于遥感影像在城市规划^[1]、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和建设“数字化城市”具有重要的应用价值;另一方面,遥感影像中纹理和特征丰富,建筑物的检测和分割提取方法种类繁多,过程具有很大相似性和通用性,可以为其他图像的目标检测和语义分割提供借鉴,因此建筑物提取方法的研究具有重要的科学意义.

然而,遥感影像含有丰富的地物信息,建筑物在遥感影像中多以面目标的形式存在,城区和乡村绿化区域较多^[2],绿色植被对于建筑物的覆盖和干扰严重,且建筑物类型和形态等复杂多样,这些均对建筑物提取带来了极大的挑战.

近年来,研究者尝试使用多种方法来实现遥感影像中建筑物的自动提取,遥感影像的大规模使用,对遥感信息和地物信息的处理与分析提出了更多的要求和挑战.传统基于手动提取特征的方法在需要人工动态调整各种参数和阈值,方法的鲁棒性和泛化能力较差.利用深度卷积神经网络可以从大量数据中自动学习特征,具有高效的特征表达能力.然而基于深度卷积神经网络的方法,也存在着计算效率低,提取结果的准确率、边界不完整等问题.

传统Mask-RCNN网络基于TensorFlow作为后端的Keras深度学习框架搭建而成,由于框架的制约,网络的性能不能够很好地发挥出来.基于PyTorch框架搭建网络,其体现出的优势不仅仅是PyTorch相比于以TensorFlow作为后端的Keras框架的优势,更表现为Mask-RCNN网络在全新的框架下性能的提升.计算机显存资源利用效率更高,计算的速度和精确度也得到了显著提升.而且新的框架不仅便于调试,高度模块化,搭建模型十分方便,数据参数在CPU与GPU之间迁移十分灵活.

针对遥感影像提取建筑物现有相关的研究成果的缺陷和不足,本文提出了一种采用全新开源的PyTorch深度学习框架和基于Mask-RCNN改进的网络结构模型的实例分割算法,更加高效与鲁棒地实现了建筑物的自动精确分割和提取.

2 算法原理与网络调整

2.1 网络结构

本文采用的深度学习分割算法基于由He KM于2017年提出Mask-RCNN实例分割算法.算法网络来源于Faster-RCNN和全卷积网络(Fully-Convolutional Network, FCN),作者在此基础上新增了改进措施.在Faster-RCNN网络的基础上新增了一个任务分支完成目标物体的像素实例分割任务,网络输出最终包含了3个任务分支,分别是目标的分类、目标框的坐标及目标区域的二值掩码实例分割,算法结构流程图如图1所示.

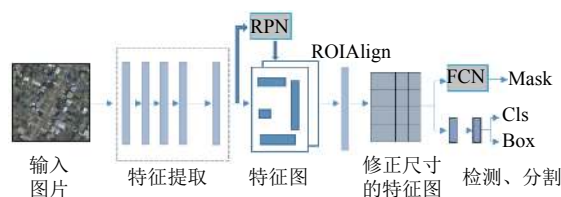


图1 算法结构流程图

图1中分类和回归部分都是由基于区域的目标检测网络Faster-RCNN^[3]完成,目标区域的像素分割由FCN^[4]完成.从图1可以看出,对于输入的图片,首先进行的是将图片输入到残差网络(ResNet)中进行特征提取,得到的主干特征图(Feature Map)通过区域生成网络(Region Proposal Network, RPN)提取可能存在的目标区域(Region Of Interest, ROI).这些ROI经过ROIAlign层被映射成固定维数的特征向量,其中两个分支经过全连接层进行分类和目标框的回归^[5],另一个分支经过全卷积运算进行上采样得到了分割区域的图片.

算法网络训练采用的多任务损失函数,通过学习减小损失函数的值,直到获得全局最优解.损失函数包含分类误差、目标框回归误差和掩码误差.

2.2 Mask-RCNN的调整和优化

传统的Mask-RCNN网络功能强大,网络的主要模块RPN存在着运算量大,效率不高的问题,FPN网

络特征图中的最高层级特征和较低层级特征之间的路径太长,不利于信息的有效融合;Mask预测仅在单个视野上执行,无法获得更加多样化的信息,针对于特殊目标的检测和分割效果欠佳.本文在继承Mask-RCNN优点的同时,根据遥感影像中建筑物的特殊性,对网络做了以下改进.

2.2.1 RPN网络的改进

RPN网络采用9种面积尺寸各异的目标框来估算原始目标在每个滑动窗口上位置上的大小,本文考虑训练集遥感影像尺寸大小,分别修改为 64×64 、 128×128 、 256×256 ,长宽比皆为1:1.调整后能够提高小目标的检测和分割效果,提高目标框的检出率.

针对RPN网络采用 3×3 的固定滑动窗口扫描特征图提取候选框带来重复的估计目标和计算的问题,本文提出了一种改进的RPN网络,结构如图2所示.将固定的滑动窗口修改为固定尺寸分割,分割后每一个窗口同样经过不同尺寸的候选框提取出ROI,之后的处理和原RPN网络相同.由于采用固定尺寸分割,改进后的RPN大大减少了估计目标区域,多种分割方法相当于构建了多种尺度的RPN,计算的效率更高,使用更少的内存空间.

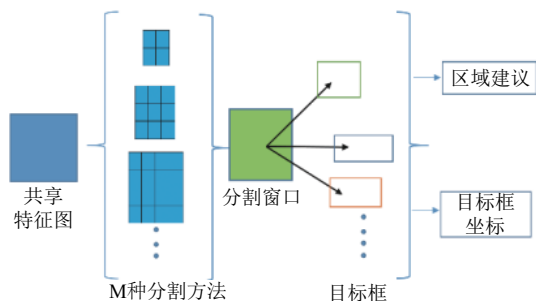


图2 改进的RPN运行机制

2.2.2 FPN网络的改进

特征金字塔网络 (Feature Pyramid Networks, FPN) 的提出是为了更好地实现顶层和底层特征图的融合和各个阶段的特征有效利用,Mask-RCNN中使用自顶向下的方式将高层级的特征与低层级的特征混合,从而提高FPN中所有具有合理分类能力的特征.

本文使用自底而上的思想进一步融合特征^[6],使得高层级特征获得低层级特征的路径变短,如图3(a)到图3(b)的虚线所示,创建自底向上的增强路径,用于缩短信息路径,改进的FPN网络利用自底向上特征中存储的精确定位信号,提升特征金字塔架构.

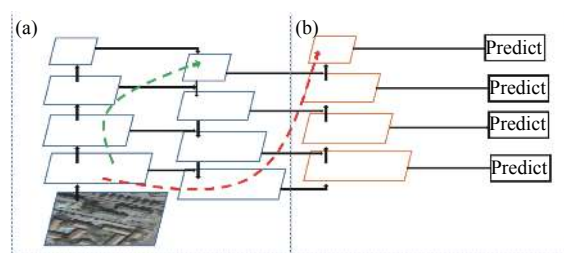


图3 改进的FPN网络结构

2.2.3 新增微全连接层支路和掩模mask的改进

在建筑物提取试验中,对于小型的建筑物,Mask-RCNN网络的提取结果较好;而在提取大建筑物时,大建筑物由于缩放比例过高而损失了大量细节信息,掩模mask的边界存在不太完整和精确之处.本文将原始网络中mask大小 14×14 像素与 28×28 像素分别修改为 40×40 与 80×80 像素.

Mask-RCNN网络中负责预测mask的组件是一个轻量级、易实现的分支,mask分支的输入是每个候选区域融合后的池化特征网格,如图4所示,主分支(传统的FCN支路)是4个连续的卷积层和一个反卷积层.其中每个卷积层核大小为 3×3 通道为256个,后面再接一个上采样2倍的反卷积层,用于预测每个类别mask的二进制像素值.

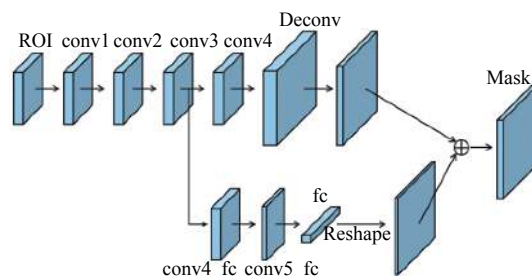


图4 mask组件与全连接层融合

本文使用一条支路从conv3连接到全连接层,中间过两个 3×3 大小卷积层conv4_fc,conv5_fc.Conv5_fc卷积层通道数减半以减少计算量,mask大小设置为 40×40 和 80×80 .全连接层产生的向量再重塑成和FCN预测的mask同样的空间尺寸.最终与FCN的输出相加得到最终预测.全连接层用于预测类不可知的背景或前景mask,不仅效率高,而且允许更多样本训练全连接层的参数,泛化能力更强.

为了捕捉每个提议的不同视图,用微小的全连接层来增加掩模预测,全连接层与最初使用的FCN具有

互补的特性,通过融合这两种观点的预测,信息的多样性会增加,产生质量更好的掩模效果。

2.3 基于全新的 PyTorch 框架搭建网络

PyTorch 是由 Facebook 于 2017 年初首次推出开源的神经网络框架,Mask-RCNN Benchmark 是一个完全由 PyTorch 1.0 写成,快速、模块化的 Faster R-CNN 和 Mask-RCNN 组件。该组件旨在让用户更容易地创建一种模型,实现对图片中目标的识别与分割。利用 Mask-RCNN Benchmark 开源项目,结合经过预处理的 IAILD 数据集,通过监督学习和迁移学习的方式,结合网络的预训练权重训练网络,保存最终的训练权重,利用预测函数预测和分割遥感影像中建筑物,实现遥感影像中的建筑物提取。

3 建筑物自动提取流程

3.1 总体流程

本文的建筑物提取流程如图 5 所示,具体步骤为: (1) 获取遥感影像; (2) 对遥感影像进行预处理; (3) 对预处理后遥感影像进行拓展、裁剪,得到大小为 512×512 的 3 通道子图像; (4) 裁剪得到的遥感影像训练集和验证集随机分成 4 个子数据集; (5) 利用标注工具 labelme 标注训练集和验证集中的遥感影像; (6) 设置网络参数、调整网络配置,如初始化学率,Epoch 数目等; (7) 载入预训练权重开始训练网络,保存训练得到的权重结果; (8) 载入训练得到的结果权重,读取测试集图片得到二值掩模图片; (9) 将二值掩模图片和建筑物标记二值图像进行像素的颜色处理; (10) 对处理结果进行性能评估和对比。

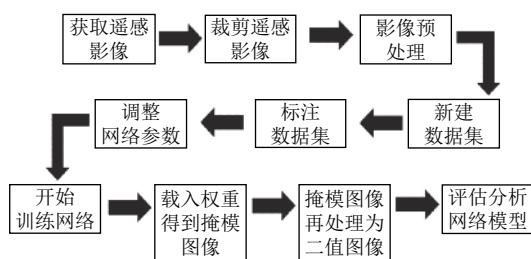


图 5 遥感影像建筑物提取流程

3.2 遥感影像中建筑物自动提取的两个阶段

Mask-RCNN 是一种多任务的深度神经网络,提取建筑物的过程分为两个阶段:第 1 阶段,Mask-RCNN 主干网络(即 ResNet101 和 FPN)提取遥感影像的特征图,随后使用 RPN 生成目标的候选框,并对候选框筛

选得到感兴趣区域。第 2 阶段,对每个感兴趣区域做类别的分类和预测、目标框的回归和目标建筑物的二值掩膜分割。

3.3 Mask-RCNN 算法步骤

(1) 网络读取一幅经过预处理后的图片;将其输入到一个预训练好的主干网络(即 ResNet101 和 FPN)中,获得对应的图像特征图的集合。(2) 对特征图集中的每一点设定预定数量的 ROI,最后获得多个最佳的候选 ROI。(3) 将候选的 ROI 送入 RPN 网络进行前景或背景的二值分类和边框回归,过滤一部分候选的 ROI。(4) 对剩余 ROI 进行 ROIAlign 操作,即将原图和特征图的像素对应起来,随后将特征图和固定的特征对应起来。(5) 最后,对 ROI 进行 N 类别的分类、位置的预测,目标框回归。通过改进的分割网络进行 mask 的预测和语义的分割,得到二值的掩模图像。具体步骤如图 6。

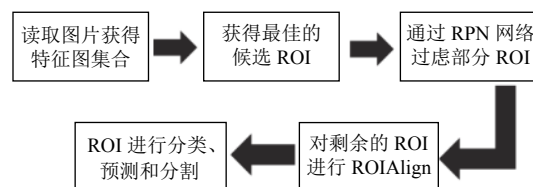


图 6 改进的 Mask-RCNN 算法步骤

4 数据集的制作及预处理

4.1 数据源与数据集

本文实验采用的遥感影像数据源为法国国家信息与自动化研究所提供的 Inria 航空影像标签公开数据集(Inria Aerial Image Labeling Dataset, IAILD 数据集)。该数据集包含 180 张 5000×5000 像素,空间分辨率为 0.3 米的航空正射彩色图像,数据集中每张图片的覆盖面积大约是 405 平方公里。这些遥感影像覆盖了五个不同的城市居民点,从人口稠密地区(如美国旧金山的金融区)到高山城镇(如美国利恩茨和奥地利蒂罗尔小镇)。

在神经网络训练的过程中,往往对输入图像的大小需要调整,首先是为了满足拥有固定节点数的全连接层的神经网络的输入要求,其次对于不同规则大的图片,为了使输入的图片不进行压缩,减轻过高的缩放比例带来的信息损失;而且如果图片尺寸太大,直接使用会造成计算速度缓慢甚至显卡内存溢出等问题。

本文把数据集随机划分为 3 部分:训练集(150

张)、验证集(5张)、测试集(25张)。为了契合神经网络的需要、遥感影像信息和计算机内存资源的有效利用,统一数据集中遥感影像的大小,将5000×5000像素影像使用黑色像素在四周边缘填充60像素,填充后遥感影像样本如图7所示,影像大小拓展为5120×5120。随后利用大小为512×512、步长为512的滑动窗口,将影像裁剪为10行10列,得到100张512×512大小的3通道遥感影像。将数据集中的所有影像填充后裁剪,最终,得到了训练集(15000张)、验证集(500张)、测试集(2500张),裁剪后的训练集和测试集中包含不同传感器不同城市,不同风格、用途、尺度的遥感影像建筑物样本,样本的增加既增强了分割模型的学习泛化能力,也缓解了数据样本的重复学习产生的过拟合问题。样本示例如图8、图9所示。



图7 原始遥感影像四周边缘填充60黑色像素

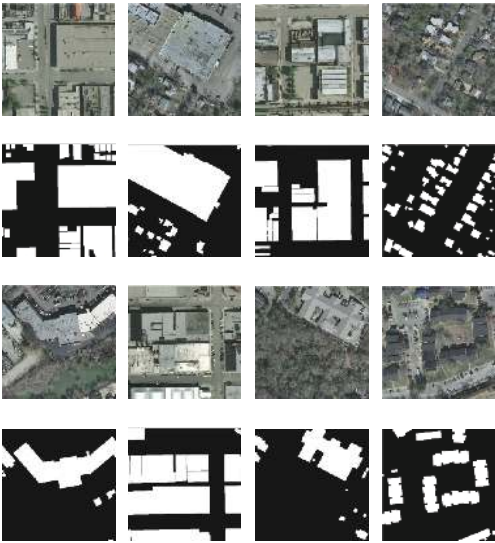


图8 不同风格、用途、尺度的建筑物样本

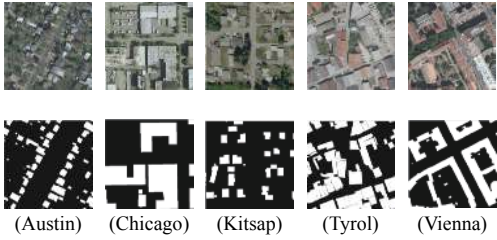


图9 数据集中不同传感器不同城市的遥感影像建筑物样本

为了更好地进行网络的训练和效果的调试,将训练集和测试集影像分为4个子数据集,各个数据集的影像均为处理后的512×512大小的3通道遥感影像,其训练集、验证集影像数量如表1所示。实验过程中可以选择4个不同量级的数据集,既可用于评估模型的准确率,还可验证本文方法的鲁棒性和泛化能力。

表1 数据集说明

数据	训练集(张)	验证集(张)	图像大小
数据集1	500	20	512×512
数据集2	1500	60	
数据集3	5000	180	
数据集4	8000	240	

4.2 遥感影像预处理

由于卫星的飞行速度、轨道高度、传感器的瞬时视角等原因^[7],在将数据进行建模前,必须先对其进行影像的预处理。

4.2.1 高斯滤波

遥感影像噪声的来源主要有两个方面,在采集过程中,图像传感器CMOS在采集图像中受传感器材料属性、电子器件的结构和工作环境的影响引入的各种噪声^[8];另外在传输过程中也会受到多种噪声污染。本文采用空间域滤波中的高斯滤波,可以平衡对图像的噪声的抑制和对图像的模糊。

4.2.2 直方图均衡化

本文采用了限制对比度的自适应直方图均衡化技术,该技术将图像分割为小的区域后进行直方图均衡化,并对局部对比度进行了限幅,这个特性也应用到全局直方图均衡化中。该技术克服了自适应直方图均衡化噪声过度放大的问题,使得图像细节更加清晰。

4.2.3 影像增强

本文采用的影像增强技术包括邻域增强和主成分分析等算法,邻域增强能够用于去噪、图像平滑、锐化等相关操作。主成分分析可用于消除特征中向量中

各个特征之间的相关性,进行特征的选择,还可用来进行高光谱图像数据的压缩和信息融合,训练影像的增强,缓解了模型的过拟合风险^[9]。

5 实验与算法评价

5.1 实验过程

本文实验使用开源的 PyTorch 学习框架,使用 Python 语言编程实现算法网络,硬件环境为联想 Think Station P900 工作站,配备有 NVIDIA GeForce GTX TITAN X 显卡 (12GB)、64 位 Ubuntu16.04 操作系统。

本文采用迁移学习方法,将 COCO2014 数据集训练得到的权重模型作为本文建筑物自动提取算法模型的预训练模型,在此预训练模型的基础上训练遥感数据集^[8]。迁移学习不仅减少了训练数据且提升了训练效率,迁移学习的模型泛化能力更好,有效提升检测模型的整体检测精度和模型性能。

本文选取数据集 1 进行学习训练,将 epoch 数值设置为 5, batch size 值设置为 8, 每个 epoch 设置步数为 10 000, 利用 Adam 算法进行网络优化, 学习率设置为 0.000 01, 所有参数根据正太分布进行初始化。整个过程在训练大概 20 小时后损失函数不再下降, 停留在 1.070 左右。

5.2 评价指标

为了定量评价遥感影像提取算法的综合性能, 本文分别采用式 (1)~式 (4) 的平均查全率 (*mRecall*)、平均查准率 (*mPrecision*)、平均准确率 (*mAP*) 和 *F1*^[10] 分数对测试集影像进行评价。*k* 为随机选取的 *k* 组测试影像, *P_{ij}* 为测试的准确率, *P_{ii}* 为正确提取的准确率。

$$mRecall = \frac{1}{k+1} \sum_{i=0}^k \left(\frac{TP}{TP+FP} \right) \times 100\% \quad (1)$$

$$mPrecision = \frac{1}{k+1} \sum_{i=0}^k \left(\frac{TP}{TP+FN} \right) \times 100\% \quad (2)$$

$$mAP = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij}} \times 100\% \quad (3)$$

$$F1 = \frac{2 \times mPrecision \times mRecall}{Precision + Recall} \times 100\% \quad (4)$$

TP (True Positive) 代表本文算法和人工标注法均提取到建筑物。*FN* (False Negative) 代表本文算法未提

取到建筑物而人工标注法能提取到建筑物。*FP* (False Positive) 代表本文算法能提取到建筑物而人工标注法未提取到建筑物。

为了验证本文方法与主流提取算法有着较大的优势, 使用基于 SVM^[11]、FCN^[12]、U-net^[13]、Mask-RCNN^[14] 等建筑物提取方法对 IAILD 数据集进行了测试。本文也测试了几个开源的数据集以验证本文方法在不同开源数据集中的泛化能力和鲁棒性。

由表 2 可知, 相比于 SVM、FCN、U-net 和 Mask-RCNN 算法, 本文算法的是个评价指标在数值上均有 5% 以上的提高, 表明本文方法的总体效果更佳, 性能优异, 有效性、准确性和鲁棒性明显提高。

表 2 在 IAILD 数据集上各种算法的比较 (%)

方法	SVM	FCN	U-net	Mask	改进Mask
<i>mAP</i>	70.75	75.34	79.22	82.23	87.31
<i>mRecall</i>	71.25	73.68	79.25	83.59	90.06
<i>mPrecision</i>	75.23	78.61	83.74	85.61	91.25
<i>F1</i>	71.39	76.07	81.43	84.59	90.66

与本文的数据集相比, Massachusetts 数据集的质量和分辨率要低很多, 而且存在着很多错误标签, 其中的平均查全率和平均查准率比本文数据集相差甚远, 这对神经网络在精确分割提取建筑物产生了负面影响。WHU 和 DOTA 是两个大场景、高分辨率的遥感建筑物数据集, 数据集中包含不同用途、不同色彩、不同大小的建筑物, 这为神经网络的训练提供了丰富的数据来源。综合对比来看 (表 3), 4 个评价指标的高分数说明本文方法的性能优异, 精确性、鲁棒性和网络的泛化性能均有不俗的表现。

表 3 在不同开源数据集中的性能表现 (%)

指标	WHU	Massachusetts	DOTA
<i>mAP</i>	86.56	78.23	85.75
<i>mRecall</i>	96.20	77.31	83.91
<i>mPrecision</i>	86.71	80.36	84.50
<i>F1</i>	91.21	78.81	84.20

5.3 实验结果分析

图 10 随机展示了 5 组利用本文方法对 IAILD 数据集中奥斯汀和芝加哥地区建筑物的提取结果, 其中图 10 中第 1、2、4 为奥斯汀地区遥感影像, 第 3 和 5 为芝加哥地区遥感影像。(a) 为原始影像, (b) 为对应的建筑物标记, (c) 为建筑物提取结果掩模覆盖图, (d) 为本文建筑物提取结果, (e) 为 (b) 和 (d) 的对比结果, 可

以发现各个类型的建筑物均大部分可以很好地提取出来. 其中使用深蓝色表示 *TP* (True Positive), 即预测结果为建筑物, 实际样本也为建筑物; 使用红色表示 *FP* (False Positive), 即实际样本为非建筑物, 但预测结果为建筑物; 用绿色表示 *FN* (False Negative), 即实际样本为建筑物, 预测结果为非建筑物; 最终将所有建筑物的预测结果用分类图来表示.

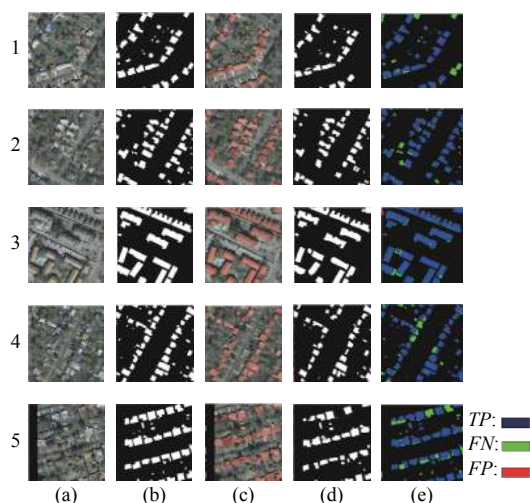


图 10 奥斯汀和芝加哥地区建筑物提取结果

本文测试所用的 IAILD 测试集试验区中建筑物有五种, 随机选取奥斯汀地区和芝加哥地区建筑物试验, 两个地区选择的背景是基于奥斯汀郊区和芝加哥城区, 从两个不同地区的不同建筑物大小分析对比本文方法的有效性和精确性.

以图 10 奥斯汀地区为例, 数据集中的奥斯汀地区主要以郊区为主, 郊区植被茂盛建筑物与周围环境信息混淆, 建筑物分散且以中小型建筑物为主, 植被覆盖对于建筑物的影响较大, 从图 10(b)(d)(e) 对比可知, 本文方法对中型建筑物提取较好, 对于小型建筑物也能很好地提取出来, 但对被植被覆盖的建筑物提取效果欠佳. 芝加哥城区建筑物覆盖密集且相对集中, 主要以大型建筑物为主, 植被覆盖影响较小, 对比发现可知, 本文方法对密集型大型建筑物均能很好的提取出来. 为了更加直观地观察提取结果对比, 如图 11 所示, 特选取 1 张尺寸为 1024×1024 的对比效果图.

综合图 10 和图 11 及所有测试集图片结果分析, 本文方法具有较好的网络的泛化性能和鲁棒性, 对 IAILD 数据集中的不同地区、不同地域、不同类型的建筑物均能取得很好的提取效果; 但对于数据集中大型建筑

物和植被覆盖的建筑物提取时也有着空洞现象, 边缘毛糙, 漏检问题依然存在. 分析主要原因为数据集影像数量的不足、迭代次数设置影响、网络模型的缺陷和数据集标记的误差等, 在以后的研究中要着力解决这些问题.

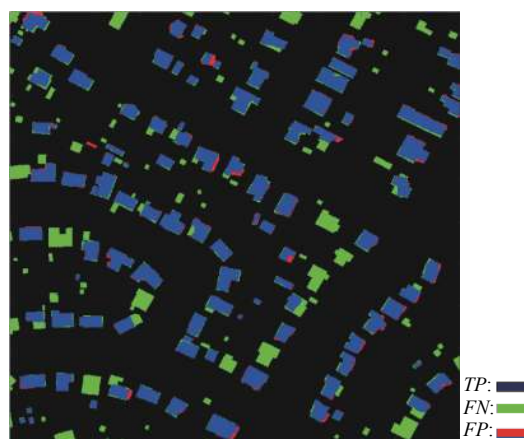


图 11 大尺寸奥斯汀建筑物提取结果样张

6 总结

本文将遥感影像中建筑物提取与深度学习相结合, 提出了一种基于 PyTorch 框架改进的 Mask-RCNN 网络提取方法, 为建筑物提取提供了一种有效的新方案. 研究表明, 本文提出的方法对于复杂类型的建筑物不仅能高效准确快捷地提取, 而且提取结果的 mAP 等评价指标均有较高的得分; 和其它算法的提取结果进行对比, 本文方法的精确性、鲁棒性和泛化能力均优于对比算法. 实现本文方法参数的最佳设置、训练影像标记的规范化以增强边缘提取效果、使其适应较大地理区域内中大型建筑物的提取、支持城市规划等相关业务是后续的研究目标.

参考文献

- 1 付发, 未建英, 张丽娜. 基于卷积网络的遥感图像建筑物提取技术研究. 软件工程, 2018, 21(6): 4-7.
- 2 邵鹏. 基于多尺度分割的高分遥感影像主要地物信息提取技术研究 [硕士学位论文]. 长春: 吉林大学, 2015.
- 3 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. 2015. 1512-1516.
- 4 Long J, Shelhamer E, Darrell T, *et al.* Fully convolutional

- networks for semantic segmentation. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 1411–1445.
- 5 欧攀, 路奎, 张正, 等. 基于 Mask RCNN 的目标识别与空间定位. 计算机测量与控制, 2019, 27(6): 172–176.
- 6 Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 8759–8768.
- 7 万里红, 杨武年, 李天华, 等. 浅谈 QuickBird 遥感卫星影像几何精校正. 测绘与空间地理信息, 2007, 30(2): 12–15, 19. [doi: [10.3969/j.issn.1672-5867.2007.02.004](https://doi.org/10.3969/j.issn.1672-5867.2007.02.004)]
- 8 李大军, 何维龙, 郭丙轩, 等. 基于 Mask-RCNN 的建筑物目标检测算法. 测绘科学, 2019, 44(10): 172–180.
- 9 唐文博. 基于卷积神经网络的高分辨率多光谱遥感图像上的城区建筑物变化检测技术 [硕士学位论文]. 杭州: 浙江大学, 2019.
- 10 Ok AO. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. ISPRS Journal of Photogrammetry and Remote Sensing, 2013, 86: 21–40. [doi: [10.1016/j.isprsjprs.2013.09.004](https://doi.org/10.1016/j.isprsjprs.2013.09.004)]
- 11 赵春晖, 李思聪. 基于多特征融合的 SVM 高分辨率遥感影像建筑物提取方法. 沈阳大学学报 (自然科学版), 2017, 29(4): 311–316.
- 12 刘文涛, 李世华, 覃驭楚. 基于全卷积神经网络的建筑物屋顶自动提取. 地球信息科学学报, 2018, 20(11): 1562–1570. [doi: [10.12082/dqxxkx.2018.180159](https://doi.org/10.12082/dqxxkx.2018.180159)]
- 13 苏健民, 杨岚心, 景维鹏. 基于 U-Net 的高分辨率遥感图像语义分割方法. 计算机工程与应用, 2019, 55(7): 207–213. [doi: [10.3778/j.issn.1002-8331.1806-0024](https://doi.org/10.3778/j.issn.1002-8331.1806-0024)]
- 14 季顺平, 魏世清. 遥感影像建筑物提取的卷积神经网络与开源数据集方法. 测绘学报, 2019, 48(4): 448–459. [doi: [10.11947/j.AGCS.2019.20180206](https://doi.org/10.11947/j.AGCS.2019.20180206)]