

基于深层卷积残差网络的航拍图建筑物精确分割方法



许华杰^{1,2} 张晨强¹ 苏国韶³

1 广西大学计算机与电子信息学院 南宁 530004

2 广西多媒体通信与网络技术重点实验室 南宁 530004

3 广西大学土木建筑工程学院 南宁 530004

摘要 针对建筑物 3D 建模场景下所需的建筑物主体轮廓俯视平面图获取成本较高、航拍图建筑物的分割精度低、建筑物屋顶存在干扰物影响分割等问题,文中提出了一种将 5 个点的位置表示为热图作为网络额外输入通道的基于深层残差网络的航拍图建筑物精确分割方法,该方法在航拍图建筑物的精确分割任务中取得了比较好的分割效果。实验结果表明,该方法具有比传统半自动分割方法 Grabcut 更高的分割精度和分割效率;具有比 DEXTR 方法更好的鲁棒性和抗干扰性。该方法可以为建筑物 3D 重建任务提供高精度的建筑物俯视轮廓图和建筑物顶部图片,还可以在航拍图建筑物数据集的制作过程中,作为一种准确和有效的掩码注释工具或半自动轮廓标注工具,以提高数据集的标注效率。

关键词: 卷积残差网络;图像分割;航拍图;3D 建模;热图

中图法分类号 TP391

Accurate Segmentation Method of Aerial Photography Buildings Based on Deep Convolutional Residual Network

XU Hua-jie^{1,2}, ZHANG Chen-qiang¹ and SU Guo-shao³

1 College of Computer and Electronic Information, Guangxi University, Nanning 530004, China

2 Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China

3 College of Civil Engineering and Architecture, Guangxi University, Nanning 530004, China

Abstract In order to solve the problems of high cost of obtaining the top plan view of the main outline of the building in the 3D modeling scenario, low segmentation accuracy of the aerial photography building, interference on the roof of the building, etc., a method of accurately segmenting the aerial photography building based on deep residual network is proposed, in which the positions of five points are expressed as heat maps as additional input channels of the network, and good segmentation effect is achieved in the task of accurately segmenting the aerial photography building. Experimental results show that the proposed method has higher segmentation accuracy and segmentation efficiency than the traditional semi-automatic segmentation method Grabcut. It has better robustness and anti-interference than DEXTR method. This method can provide high-precision top-view contour map and top-view picture of buildings for 3D reconstruction of buildings, and can also be used in the production process of aerial photography building data sets as an accurate and effective mask annotation tool or semi-automatic contour annotation tool to improve the annotation efficiency of datasets.

Keywords Convolutional residual network, Image segmentation, Aerial photography, 3D modeling, Heatmap

1 引言

随着社会经济的发展,城市建筑物的数量和类别也在迅速增长,数字化管理对城市建筑物 3D 重建提出了更高的要求^[1]。此外,随着无人机航拍技术的普及,城市建筑物的航拍图像也更容易获得,这些航拍图像中含有丰富的建筑物特征

信息,从航拍图像中检测和分割出建筑物目标对城市 3D 重建具有重要意义。

目前常见的图像分割方法有阈值分割方法^[2]、边缘检测方法^[3]等,但是这些方法都存在一定的局限性,在复杂的高分辨率的航拍图建筑物分割任务中难以取得精确的分割效果。而人工分割方法虽然可以获得比较精准的图像分割效果,但

到稿日期:2020-05-21 返修日期:2020-10-23 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:广西壮族自治区科技计划项目(2017AB15008);崇左市科技计划项目(FB2018001);广西高等学校高水平创新团队及卓越学者计划 This work was supported by the Science and Technology Plan Project of Guangxi Zhuang Autonomous Region (2017AB15008), Science and Technology Plan Project of Chongzuo(FB2018001) and High Level Innovation Team and Outstanding Scholar Program of Universities in Guangxi Province.

通信作者:许华杰(hjxu2009@163.com)

是分割耗时长、效率低。因此,人工参与交互的半自动图像分割方法也有着较高的研究价值和应用价值。如人工参与的交互式前景与背景分割算法 Grabcut^[4],该方法在目标物体外人工选择一个包含物体的矩形框,可以将物体和它的背景分离,达到提取物体区域的效果,也可以根据效果增加额外的手工划线(人工指定一些像素区域属于前景),以达到更好的分割效果。Grabcut 算法因其广泛的实用性在图像分割领域有着重要地位,但是其在精度上尚达不到对航拍图建筑物进行精确分割的要求。Papadopoulos 等^[5]提出了一种通过极端位置的点击来标注对象的方法,其使用上、下、左、右 4 个极端点来增强 Grabcut 对象分割效果。这些极端点属于对象的顶部、底部、最左侧和最右侧部分,通过 4 个极端点的人工标注可以获得高质量的物体边界框,而人工标注的极端点所提供的信息比边界框更多,能为对象分割提供更多有用信息。Maninis 等^[6]提出了 DEXTR(Deep Extreme Cut)方法,其使用 4 个极端点位置表示的热图,作为卷积神经网络^[7]的额外输入通道,以精确分割图像和视频,该方法为半自动的方法,将手工标注的极端点与深度残差网络结合起来,在多个场景下取得了较好的分割效果。

针对建筑物 3D 重建任务中所需要的精确建筑物图片分割需求,本文对半自动图像分割方法 DEXTR 进行了改进,提出一种基于深层残差网络的航拍图建筑物精确分割方法 E-DEXTR(Enhanced-Deep Extreme Cut)。该方法将建筑物上标注的 4 个极端点与 1 个中间位置点表示为热图作为卷积神经网络除图像的 RGB 通道之外的额外输入通道,从而实现了

对航拍图中的建筑物进行精确分割。相比其他半自动图像分割方法,E-DEXTR 方法具有较高的分割精度和较好的抗干扰性,在航拍图建筑物的精确分割任务中取得了更好的效果。

2 基于深层残差网络的航拍图建筑物精确分割方法

在工程实际的航拍图建筑物实例分割任务中,利用 DEXTR 方法得到的分割结果可能并不是一个完整的轮廓图,分割结果常常受到航拍图中干扰因素的影响。图 1 给出了工程实际航拍图分割任务中建筑物屋顶存在干扰物的一些案例,图 1(a)中屋顶上存在居民晾衣杆和蓝色洗衣盆等明显的干扰物,图 1(b)和图 1(c)中存在太阳能热水器和卫星天线等干扰物。在城镇建筑物的航拍图中,没有干扰物的建筑物屋顶是很少的,各种干扰因素普遍存在。DEXTR 方法在这些存在干扰信息的建筑物俯视图分割任务中得到的分割结果常常是不完整的俯视轮廓图,残缺的分割结果无法应用于需要精确建筑物顶部轮廓的 3D 建模任务中。

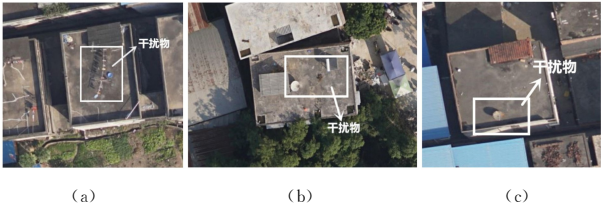


图 1 存在干扰物的航拍图建筑物

Fig. 1 Aerial view of buildings with interference

为了解决上述问题,得到完整、可用的建筑物俯视轮廓

图,本文在 DEXTR 方法的基础上,提出了改进方法 E-DEXTR,其额外添加了 1 个手工标注的位于目标建筑物对象中间的点及其他 4 个极端点的位置,将 5 个点的位置表示为热图,并与待分割图片一起作为深度残差网络的输入。实验结果表明,本文方法可应用于航拍图建筑物的精确分割。

2.1 五点热图的输入通道

E-DEXTR 方法将手工标注的 4 个极端点以及 1 个中间点的位置信息表示为热图作为额外输入通道,与航拍图 R, G, B 通道一起作为卷积残差网络的输入。对于所标注的 4 个极端点以及 1 个中间点,分别以每个点为中心生成高斯分布的概率区域,从而得到包含 5 个点位置信息的热图。所生成的热图大小与输入图片的大小相同,在每个关键点的位置处生成一个呈高斯分布的概率区域,热图矩阵中的值表示对应像素点属于需要分割的某个建筑物的概率,类似于人体姿态检测中骨骼关键点的热图^[8]。在以每个点的位置为中心生成的高斯分布概率区域中,区域的中心(为标注的极端点或者中间点)概率值最大,最接近于 1,越向四周概率越小。高斯概率分布区域中每个点所对应的值由二维高斯函数计算得到,计算公式如式(1)^[9]所示:

$$g(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{((x-x_0)^2+(y-y_0)^2)}{2\sigma^2}}$$
 (1)

其中, x_0, y_0 表示中心点的坐标; x, y 是高斯概率区域中每个像素的坐标; σ 表示方差,决定了高斯分布的有效范围。

图 2 给出了基于 4 个点与 5 个点创建的热图的可视化效果,图中对应的 $\sigma=10$ 。图 2 中,原图是采用 5 点标注的航拍图(5 点包括 4 个极端点和 1 个中间点,中间点的标注位置可以在建筑物轮廓内任意取);DEXTR 方法采用 4 个极端点位置生成热图;E-DEXTR 方法以 5 个点的位置生成热图。在热图的可视化效果图中,热图中像素值越高(颜色越亮)的区域所对应航拍图的像素点属于需要分割的建筑物的概率值就越大,因此每个中心点区域的概率值最大,接近于 1。E-DEXTR 方法在保留 DEXTR 方法将 4 个极端点作为输入的基础上,通过再额外增加 1 个中间点,创建了含有 5 个点位置信息的热图,从而提供了更多的像素点分类指导信息。另外,针对屋顶存在阴影或者太阳能热水器等干扰物的情况,该方法可以将中间点的位置选择标注在干扰物上以获取干扰物所在区域的位置特征信息,用于实现网络对建筑物屋顶的精确分割。

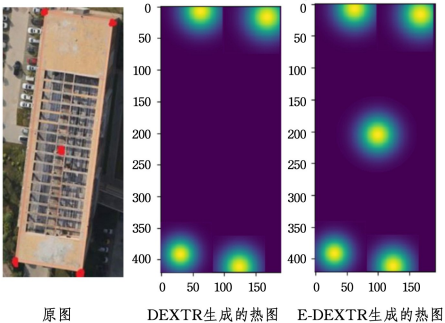


图 2 两种方法生成的热图的可视化效果图

Fig. 2 Visualization of heat maps generated by two methods

为了更好地提取分割对象的特征,输入图像首先会对通

过 4 个极端点得到的矩形边界框进行裁剪,得到的裁剪区域包含整个建筑物对象。为了使裁剪部分中包含更多的背景信息,可以将矩形边界框适当放大,在本文实验中,矩形边界框向外放大 10 个像素的距离。包含 5 个点位置信息的热图的裁剪位置和输入图像的裁剪位置一致。在对 4 个极端点以及 1 个中间点的位置进行标注之后,得到特征提取网络的输入,其包括含有目标对象的裁剪区域(分成 R,G,B 3 个通道)和该裁剪区域所对应的包含 5 个点信息的热图(热图通道)。

2.2 ResNet-101 和金字塔场景解析模块

E-DEXTR 方法采用 He 等^[10]提出的深度卷积残差网络 ResNet-101 作为特征提取网络,构成了整个网络架构的骨干部分。深度卷积残差网络有多个不同层数的版本,比较常见的有 18 层、34 层、50 层、101 层和 152 层,本文选择 101 层的版本是因为它已被证明在各种图像分割场景下都能取得较好的特征提取效果^[11]。所采用的 ResNet-101 网络结构如图 3(a)所示。

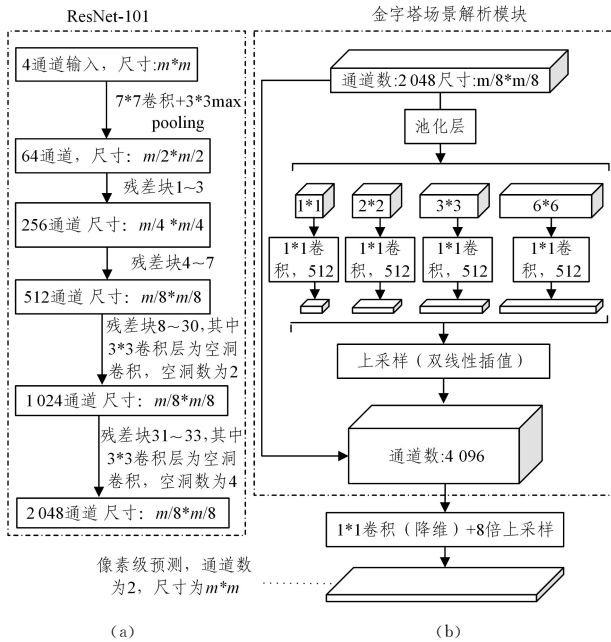


图 3 ResNet-101 网络和金字塔场景解析模块结构示意图

Fig. 3 Structure of ResNet-101 and pyramid scene analysis module

从图 3 可以看到,网络的输入通道数为 4,包括图像的 R,G,B 通道以及 5 个标注点生成的热图通道。从输入端开始,网络的第一层为卷积核大小为 7×7 的卷积层,之后连接了一个池化核大小为 3×3 的最大池化层,池化层之后连接了 33 个残差块,残差块连接的使用有助于解决深层网络中出现的梯度消失问题^[12]。为了输出较高分辨率的特征图以进行像素级的预测,移除 ResNet-101 最后的池化层和全连接层,并且取消第 8 个和第 31 个残差块中的下采样操作(将残差块中第一层的 1×1 卷积操作的步长由 2 改为 1)。另外,为了扩大特征图的感受野(感受野越大越有利于语义分割),分别在残差块 8~30 和 31~33 中引入 3×3 的空洞卷积层(也叫膨胀卷积层)^[13],空洞数分别为 2 和 4。最后 ResNet-101 输出的特征图尺寸为输入的 $1/8$ 。在 ResNet-101 之后连接金字塔场景解析模块^[14],目的是将全局上下文信息聚合到最终的

特征图中,以提升网络的预测精度。所采用的金字塔场景解析模块的网络结构如图 3(b)所示。从图中可以看到,金字塔场景解析模块将 ResNet-101 输出的特征图作为输入,使用 4 个不同尺度的池化核对输入特征图进行池化操作,得到 4 个不同尺度的池化后的特征图。池化后的特征图分别通过卷积核大小为 1×1 、步长为 1 的卷积层将通道数降低为输入特征图的 $1/4$ 。再将这 4 个低维特征图利用双线性插值法进行上采样,得到和输入特征图一样的尺度大小。最后将这 4 个不同尺度的特征图和输入特征图连接起来,得到最终特征图。最终特征图再通过卷积核大小为 1×1 、步长为 1 的卷积层降低通道数,并且经过上采样操作(双线性插值)得到与 ResNet-101 的输入图像相同的尺寸。

ResNet-101 网络使用的激活函数统一为 ReLU 函数,函数表达式如下:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (2)$$

其中, $f(x)$ 表示激活层的输出, x 表示输入。

E-DEXTR 方法网络架构中采用 Softmax 分类器对由 4 个极端点裁剪得到区域中的像素进行预测,其使用的 Softmax 函数的表达式如下:

$$p(x_i) = \frac{\exp(x_i)}{\sum_{h=1}^k \exp(x_h)} \quad (3)$$

其中, x_h 表示网络最后一层的输出, Softmax 函数的输出是由 x_i 计算得到的概率值表示。在本文的航拍图建筑物精确实例分割任务中,式(3)中 $k=2$ 且 $\sum_{i=1}^k p(x_i) = 1$ 。网络的输出表示相应像素是否属于待分割对象(建筑物)的概率值。将 Softmax 函数的输出结果作为损失函数的输入,损失函数用于评估模型的预测值与真实值之间的差异程度,作为网络优化的目标函数。E-DEXTR 方法中损失函数采用的是标准交叉熵损失函数,其表达式如式(4)^[6]所示:

$$L = \sum_{j \in Y} C(y_j, \hat{y}_j), j \in 1, 2, \dots, |Y| \quad (4)$$

其中:

$$C(y_j, \hat{y}_j) = \begin{cases} -\log(p), & \text{if } y = \hat{y}_j \\ -\log(1-p), & \text{otherwise} \end{cases} \quad (5)$$

在式(4)与式(5)中, y 表示单个像素, Y 表示裁剪区域图中所有像素的集合, y_j 表示像素的标签, $\hat{y}_j$ 表示预测值, $C(\cdot)$ 表示像素标签与预测值之间的标准交叉熵损失。

2.3 总体网络结构

图 4 为 E-DEXTR 方法所采用的网络结构示意图。首先将图片(含有目标对象的裁剪区域)的 R,G,B 通道以及 5 个标注点生成的热图通道输入 ResNet-101 中得到一个特征图;然后将该特征图输入金字塔场景解析模块中得到聚合全局上下文信息的最终特征图;最后根据最终特征图得到分割结果。使用在 ImageNet 数据集上预训练得到的参数作为网络的初始化权重参数,其已被证实对各种图像处理任务是有益的^[15],可以实现训练过程的加速,因此 DEXTR 方法在图像分割的实验中使用在 ImageNet 数据集上预训练的模型权重作为网络的初始化参数,然后在 PASCAL2012 数据集上进行训练。E-DEXTR 方法采用与 DEXTR 方法相同的训练策略。

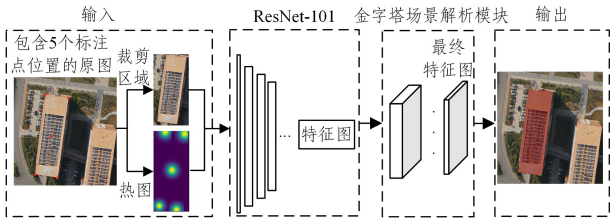


图 4 E-DEXTR 的网络结构示意图

Fig. 4 Schematic diagram of E-DEXTR model

3 实验与结果

3.1 实验平台和评估标准

为了测试本文提出的基于深层残差网络的航拍图建筑物精确分割方法 E-DEXTR 在航拍图建筑物分割任务中的效果及性能,我们在分割效果及精度、抗干扰性方面做了一系列实验,对比方法选择与本文方法定位、用途类似的 Grabcut 方法和 DEXTR 方法。实验平台的硬件配置为单 CPU Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz,单 GPU NVIDIA GTX 1080Ti;操作系统为 Windows 10(64 位);开发环境为深度学习平台 PyTorch 1.0。

实验所采用的数据集是利用无人机航拍的南宁市部分地区的俯瞰图。该数据集包括 300 张图片,从中随机选出 200 张作为训练数据集,从剩余的 100 张中随机选出 40 张作为普通测试数据集,再从剩余 60 张图像中人工挑选出屋顶含有明显干扰物的建筑物图像 10 张,作为干扰测试数据集。数据集标注工具使用的是 Labelme 软件,训练数据集和测试数据集中建筑物的轮廓由人工勾勒得到,人工选择极端点和中间点的位置并在 Labelme 软件中标注轮廓区域和点的标签信息。在采用 E-DEXTR 方法和 DEXTR 方法进行训练时,都采用 DEXTR 网络在 PASCAL2012 数据集上训练得到的模型权重作为网络的初始参数,之后通过训练数据集进行训练得到所需的航拍图建筑物精确分割模型。

实验采用的评估标准是交并比 (Intersection-over-Union, IoU),它是指模型所预测目标区域与标注真实区域之间的交并比,常用作目标检测领域和图像语义分割领域的度量标准。IoU 的计算方法如下:

$$IoU = \frac{A \cap B}{A \cup B} \tag{6}$$

其中,A 代表真实物体区域,B 代表模型预测出的物体区域。 $0 \leq IoU \leq 1$,IoU 值越大,表示分割结果越精准。

3.2 分割效果与精度测试

首先测试 E-DEXTR 方法在航拍图建筑物分割任务中的效果。E-DEXTR 和 DEXTR 两种方法在训练时使用相同的训练数据集,采用相同的超参数和初始权重,测试时使用相同的测试数据集,用于 E-DEXTR 方法的训练数据集和测试数据集的标注是在 DEXTR 方法所使用的相应数据集标注(四极端点标注)的基础上添加 1 个中间点标注。Grabcut 方法在实验中使用的是 Python 下 OpenCV 环境中提供的函数包。为了降低随机误差对实验结果造成的影响,在对测试数据集进行测试的实验过程中,每组实验分别在相同条件下重复进行 10 次,最终结果取 10 次实验的平均值。

图 5 为训练好的 E-DEXTR 模型在测试图片上的分割效果示例图,从图中分割出来的建筑物可以看到,用户只需要在想要进行分割的目标建筑物顶部的上、下、左、右位置各标上一个点,就可以得到该建筑物完整的、准确的顶部轮廓区域。E-DEXTR 方法的分割精度可以满足 3D 建模对建筑物屋顶轮廓信息的需求。



图 5 训练好的 E-DEXTR 模型分割效果示例图

Fig. 5 Segmentation effect of the trained E-DEXTR model

3 种方法在同一测试集上的实验对比结果如表 1 所列。可以看到,DEXTR 方法的分割精度比 Grabcut 方法提升了 18%,并且单个建筑物的平均分割时间大幅减少;E-DEXTR 方法的分割精度比 DEXTR 方法提高了 3.2%,单个建筑物的平均分割时间与 DEXTR 方法相当,但比 Grabcut 方法所需的分割时间缩短了 64.6%。相对于 Grabcut 方法,E-DEXTR 方法与 DEXTR 方法在分割精度和分割效率方面性能更优;E-DEXTR 方法与 DEXTR 方法相比分割效率相当,在分割精度方面占优。以上结果的原因在于两种方法输入的人工标注点的信息不同,在 4 个极端点位置相同的情况下,E-DEXTR 方法比 DEXTR 方法多出 1 个位于建筑物对象中部的点,由 5 个点生成的热图比 4 个点生成的热图提供了更多的特征信息,更有利于网络训练和优化,因此训练出来的模型更优,提高了模型的分割精度。

表 1 3 种方法在普通数据集上的测试结果

Table 1 Test results of three methods on common datasets

方法	IoU/%	单个建筑物的平均 分割时间/s
Grabcut	70.1	7.0
DEXTR	88.1	1.6
E-DEXTR	91.3	1.9

3.3 抗干扰性测试

为了测试 E-DEXTR 方法的抗干扰性,分别测试了 E-DEXTR 方法与 DEXTR 方法在相同条件下训练得到的模型在有干扰物的建筑物图片上的分割效果,用于 E-DEXTR 方法的测试数据集的标注是在 DEXTR 方法所使用的测试数据集标注(四极端点标注)的基础上添加 1 个中间点标注。

图 6 是训练好的 E-DEXTR 模型在存在干扰物的建筑物图片上的分割效果图。为了使测试更全面,除了常规的四边形建筑物(见图 6(a)左图)之外,还选择了部分不规则的多边形建筑物(见图 6(a)右图)进行测试。测试过程中,E-DEXTR 方法中第 5 个点(中间点)的位置可以选择在建筑物屋顶范围内的干扰物上,为建筑物分割提供干扰物的相关信息。如图 6(a)左图的第 5 个点位于屋顶内的晾衣架区域,而图 6(a)右图的第 5 个点位于屋顶左下角的阳光阴影区域。图 6(b)是模型的分割效果图,为方便对比,其中人工标注的建筑物实

际轮廓由绿色线条标出,模型输出的分割结果由红色区域表示,建筑物分割的效果可以由绿色线条轮廓与红色分割区域之间的吻合程度直观反映出来。从分割效果可以看到,绿色线条轮廓与红色分割区域之间的吻合程度较好,模型预测的建筑物区域轮廓完整,对四边形建筑物和不规则多边形建筑物的预测都比较精确,分别包含了晾衣架区域和阳光阴影区域,分割结果更贴近实际。不管是常规的四边形建筑物还是不规则的多边形建筑物,E-DEXTR 方法在干扰性强的航拍图建筑物上都有着良好的分割效果,位于屋顶干扰物上的点为建筑物分割提供了更多与干扰物相关的特征信息,更有利于消除干扰物对建筑物分割效果的负面影响。

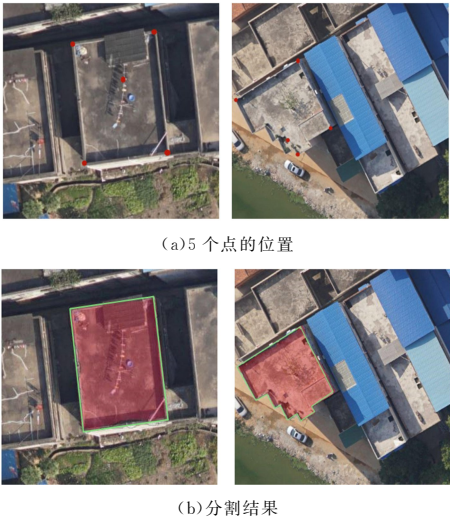


图6 存在干扰物的建筑物场景下 E-DEXTR 模型的分割效果
(电子版为彩色)

Fig. 6 Segmentation effect of E-DEXTR model in building scenes with interfering objects

表 2 为两种方法在干扰性强的航拍图数据集上的测试结果。可以看到,在干扰性强的航拍图数据集上,训练好的 E-DEXTR 模型比 DEXTR 模型的分割精度提高了 7.2%,但单个建筑物的平均分割时间也有少量增加。对于航拍图的建筑物精确分割任务来说,牺牲少量时间换取分割精度的有效提高是值得的。E-DEXTR 方法在干扰性强的航拍图建筑物上比 DEXTR 方法能获得更高的分割精度,这得益于位于屋顶干扰物上的点为建筑物分割提供了更多与干扰物相关的特征信息,从而使得建筑物分割更精准。

表 2 在干扰性强的航拍图数据集上的测试结果
Table 2 Test results on datasets of aerial image with strong interference

方法	IoU/%	单个建筑物的平均 分割时间时间/s
DEXTR	81.4	1.5
E-DEXTR	88.6	1.9

从以上两方面的实验结果可以看到,将 5 个点位置信息生成的热图作为网络额外输入通道的 E-DEXTR 方法比 DEXTR 方法的分割效果更好、精度更高以及抗干扰性更好。这说明增加 1 个中间点信息输入的改进措施的确有助于提高模型的分割精度。另外,还可以看出 E-DEXTR 方法对不规

则的建筑物的分割效果也很好。由于在对建筑物数据集进行标注的过程中,需要人工使用标注软件勾勒出建筑物的轮廓,而不规则建筑物的轮廓标注起来比较耗时,人力成本比较高,用 E-DEXTR 方法只需要标注 5 个点就可以得到目标建筑物的精确轮廓区域,因此 E-DEXTR 方法也可以作为一种准确、有效的掩码注释工具或半自动轮廓标注工具用于航拍图建筑物数据集的制作任务中,以提高数据集掩码注释或轮廓标注的效率。

结束语 针对建筑物 3D 建模场景下所需的建筑物主体轮廓的俯视平面图获取成本比较高、航拍图建筑物的分割精度不高、建筑物屋顶存在干扰物影响分割等问题,本文提出了一种将 5 个点的位置表示为热图作为网络的额外输入通道的基于深层残差网络的航拍图建筑物精确分割方法——E-DEXTR,该方法在航拍图建筑物的精确分割场景下取得了较好的分割效果。实验结果表明,与传统半自动分割方法 Grabcut 相比,E-DEXTR 方法在航拍图建筑物分割任务中具有更高的分割精度和分割效率;与 DEXTR 方法相比,E-DEXTR 方法有着更好的鲁棒性和抗干扰性。E-DEXTR 方法可以为建筑物 3D 重建任务提供高精度的建筑物俯视轮廓图和建筑物顶部图片,还可以用于航拍图建筑物数据集的制作过程中,作为一种准确和有效的掩码注释工具或半自动轮廓标注工具,提高数据集的标注效率。

本文为了获得精确的分割结果所付出的代价是需要人工交互的参与,这在对航拍图中少量建筑物进行分割的任务中是可行的。但对于大规模建筑物分割的场景,需要一种在分割过程中不需要人工参与的全自动建筑物分割方法,这将是未来的研究方向。

参 考 文 献

[1] CORNELIS N, BASTIAN L, KURT C, et al. 3D Urban Scene Modeling Integrating Recognition and[J]. International Journal of Computer Vision, 2008, 78(2/3): 121-141.

[2] SAHOO P K, SOLTANI S, WONG A K C. A survey of thresholding techniques[J]. Computer Vision Graphics & Image Processing, 1988, 41(2): 233-260.

[3] REUTER M, BIASOTTI S, GIORGI D, et al. Discrete Laplace-Beltrami operators for shape analysis and segmentation[J]. Computers & Graphics, 2009, 33(3): 381-390.

[4] ROTHER C, KOLMOGOROV V, BLAKE A. Grabcut: Interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics (TOG), ACM, 2004, 23(3): 309-314.

[5] PAPADOPOULOS D P, UIJLINGS J R R, KELLER F, et al. Extreme clicking for efficient object annotation[C]// IEEE International Conference on Computer Vision. 2017: 4930-4939.

[6] MANINIS K K, CAELLES S, PONT-TUSET J, et al. Deep extreme cut: From extreme points to object segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 616-625.

[7] KRIZHEVSKY A, SUTSKEVER L, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Sys-

tems, Curran Associates Inc. 2012;1097-1105.

[8] PAPANDREOU G,ZHU T,KANAZAWA N,et al. Towards accurate multi-person pose estimation in the wild[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;4903-4911.

[9] KHORBOTLY S,HASSAN F. A modified approximation of 2D Gaussian smoothing filters for fixed-point platforms[C]// the IEEE 43rd Southeastern Symposium on System Theory. IEEE, 2011;151-159.

[10] 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.

[11] CHEN L C,PAPANDREOU G,KOKKINOS I,et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40(4):834-848.

[12] ZHAO Z C,LUO Z,WANG P Y,et al. Survey on Image Classi-

fication Algorithms Based on Deep Residual Network[J]. Computer Systems & Applications, 2020, 29(1):14-21.

[13] YU F,KOLTUN V,FUNKHOUSER T. Dilated residual networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;472-480.

[14] ZHAO H,SHI J,QI X,et al. Pyramid scene parsing network [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;2881-2890.

[15] TAN C,SUN F,KONG T, et al. A survey on deep transfer learning[C]// International Conference on Artificial Neural Networks. Cham:Springer, 2018;270-279.



XU Hua-jie, born in 1974, Ph.D, associate professor, is a senior member of China Computer Federation. His main research interests include artificial intelligence, acoustic signal recognition and computer vision.