

基于深度卷积网络的遥感影像建筑物分割方法

余 威,龙慧云

(贵州大学 计算机科学与技术学院,贵州 贵阳 550025)

摘 要:大规模可见光遥感图像的建筑物提取是遥感图像分析领域中的一项重要工作,但是,在真实环境下,城市建筑物的尺寸范围变化大、颜色纹理轮廓复杂,加上树木等造成的遮挡,以及光照等原因,影响建筑物分割的精度。因此,文中设计并实现了两种端到端全卷积神经网络的分割方法,并在两个网络模型中加入剪裁层用以解决预测结果中产生的拼接痕迹问题,同时将 IOU 评价标准变形加入损失函数中,来提高模型分割精度。两个模型以不同尺度的遥感影像作为网络的输入,将网络模型输出结果采用加权的方式进行融合,进一步提高遥感影像建筑物识别和分割精度。在公开的 Inria 遥感影像数据集上的实验证明了该方法在遥感影像建筑物分割上的有效性。

关键词:全卷积神经网络;遥感影像;建筑物分割;模型融合

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2019)06-0057-05

doi:10.3969/j.issn.1673-629X.2019.06.012

A Building Segmentation Method Based on Deep Convolution Networks for Remote Sensing Imagery

YU Wei, LONG Hui-yun

(School of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

Abstract: Building extraction technique based on large-scale optical remote sensing images plays an important role in the field of remote sensing image analysis. But in the real environment, due to the big range of urban building's size, the complexity of building's colors, texture and contour, the occlusion of trees, as well as the illumination intensity, the precision of building segmentation is decreased. In order to improve the accuracy of building segmentation, two kinds of end-to-end full convolution networks (FCN) are proposed and realized, then crop layer is added to these two models to solve the visible boundary on predicted patches. Meanwhile IOU index are added into the loss function to improve the segmentation accuracy. These two networks use different scale images as input, and the two output images are fused in a weighted way. The experiment on Inria aerial imagery dataset shows that this method is effective in building segmentation of remote sensing imagery.

Key words: full convolution networks; remote sensing imagery; building segmentation; model fusion

1 概 述

遥感影像目标识别作为当前遥感影像应用领域中的主要研究内容,具有重要的理论研究意义和广泛的应用价值^[1]。以城市区域的高分辨率遥感影像为例,大部分识别目标是建筑物、道路和桥梁等人工地物,这些目标的有效识别和精确定位一直以来都是急需解决的问题。而这其中的遥感影像建筑物识别在民用领域和军事领域都有较为重要的意义,为灾害应急和人民生活提供便利。

近年来,传统计算机视觉技术进行高分辨率遥感影像地面建筑物识别提取主要是利用建筑物的多边形

形状信息、颜色和纹理特征^[2],随后使用传统机器学习分类算法(例如 AdaBoost、支持向量机、随机森林等)对这些特征进行分类,并完成建筑物的提取^[3-5]。但是真实环境下,城市建筑物有着复杂多变的纹理和颜色,以及会受到阴影的遮挡,这些因素使上述算法在复杂的建筑物提取场景下处理能力不足^[6]。

从遥感影像中进行建筑物识别,在计算机视觉领域中被认为是语义分割问题,语义分割是通过一定的方法,将整幅图像分割成具有一定语义含义的区域块,并识别出每个语义块的类别,实现从底层到高层语义的推断过程^[7]。最终得到一幅逐像素语义标注的分割

收稿日期:2018-07-20

修回日期:2018-11-23

网络出版时间:2019-03-06

基金项目:国家自然科学基金(61741124)

作者简介:余 威(1992-),男,硕士生,研究方向为图像分割、深度学习;龙慧云,博士,副教授,研究方向为 Web 服务的形式化分析。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190306.0952.064.html>

图像。建筑物分割是一种二分类的语义分割,分割后得到建筑物区域和非建筑物区域。

卷积神经网络(CNN)作为计算机视觉任务上成功的监督分类方法,将其应用在图像语义分割上有以下两个问题:一是输出结果相对输入分辨率降低,影响最终的分类精度;二是忽略了上下文信息。图像中物体之间存在局部上下文关系以及物体与图像之间存在全局上下文关系,CNN模型会将这些信息忽略。为了解决这些问题,Jonathan Long等提出了Fully Convolutional Networks(FCN)^[8];Zheng等^[9]将卷积神经网络的输出作为全连接条件随机场(full connected CRF)的输入,提高了语义分割准确率。此外,还有一些跳越连接(skip-connections)网络结构^[8,10-11],膨胀卷积(dilated convolutions)网络结构^[12]能在一定程度上提高整体分割准确率。虽然这些方法在处理自然图像时性能较好,但是它们并不太适合处理遥感影像。因为高分辨率遥感影像每张影像包含成百上千的建筑物目标,并且在影像中建筑物结构精细,这些都与自然图像不同。但是可以借鉴自然图像语义分割思路,结合不同网络模型优点,研究一种对复杂场景具有较强鲁棒性的遥感影像建筑物分割方法。

因此,文中构建了两种端对端的全卷积神经网络结构,并将其应用在遥感影像建筑物分割上。此外,重新设计结合交叉熵损失函数和Jaccard index的联合损失函数,用以提高遥感影像建筑物分割精度。受ImageNet竞赛集成方法^[13-14]的启发,文中提出融合两种模型的遥感影像建筑物分割方法来提高整体分割精度。

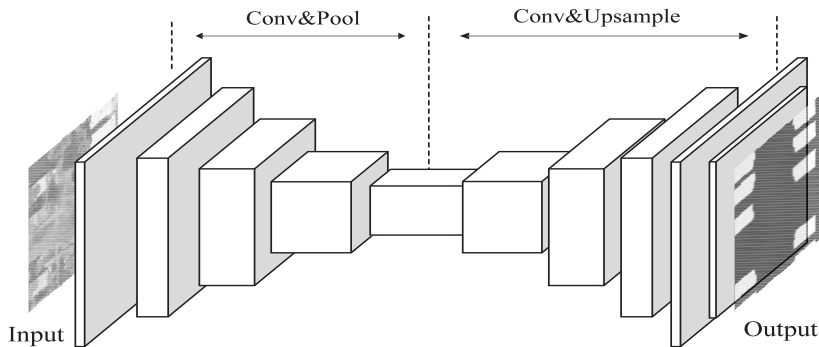


图1 编解码分割网络结构

大小为 128×128 像素的图像作为网络输入。在网络编码部分,使用VGG16结构进行特征提取,五个卷积结构部分分别由64个、128个、256个、512个和512个 3×3 的卷积核进行特征提取,紧随其后的是 2×2 的最大池化层。在解码部分,分别由五个 2×2 的上采样层通过双线性插值进行分辨率还原。根据式1,加入剪裁层,将 128×128 的特征图剪裁至 80×80 大小。在最后一层使用卷积核大小为 1×1 的卷积层。由于目

2 端对端全卷积网络模型

为了从理论上表示模型,用 S 表示遥感影像, M 表示与其对应的GT(ground truth)图像。当遥感影像 S 的 (i,j) 位置的像素表示为建筑物时,与之对应的相同位置上GT图像的 $M(i,j) = 1$ 。建立的模型的目标就是从样本数据中去学习 $p(M(i,j) | S)$ 。

在高分辨率的遥感影像中,一张图往往包含成千上万个像元,这使得图像占用很大的存储空间。由于计算机硬件的限制,只能对一整张图像进行分块预测,再对预测图进行拼接恢复。直接的预测拼接会有很明显的拼接痕迹,影响建筑物的分割精度。解决这一问题的方式是在重叠的图像块上进行预测。文中提出在模型中加入剪裁层的方式来解决这个问题,两种端对端的模型均是基于此思想。即:

$$p(N(M(i,j), w_m) | N(S(i,j), w_s)) \quad (1)$$

其中, $N(I(i,j), w)$ 表示在图像 I 上以 (i,j) 为中心,大小为 $w * w$ 的图像块。将给定大小为 $w_s * w_s$ 的遥感影像输入模型,经模型预测得到大小为 $w_m * w_m$ 的预测($w_m < w_s$)。此外,这种解决方式能降低模型的计算时间。

2.1 编解码分割网络模型(ED-Net)

典型FCN模型通过转置卷积一次性将底层特征图恢复成原图大小,致使模型分割精度不高。与FCN不同的是,编解码网络逐级通过卷积和上采样恢复图像分辨率,这样模型能保留更多细节特征。整个网络模型的结构^[10]如图1所示。

2.2 U-Net 分割网络模型

U-Net网络^[11]与编解码网络模型结构不同的是,它将包含上下文信息的低层级的特征图和包含图像细

节的高层级的特征图结合起来,实现了分割的精确定位。具体结构图和各特征图维度如图2所示。

输入网络中的是3维的RGB遥感影像,大小为 112×112 像素,通过滑动窗口选取像素块。在对图像进行卷积(卷积核大小为 3×3)后进行批量正则化(batch normalization)操作,让图像像素值的均值为0

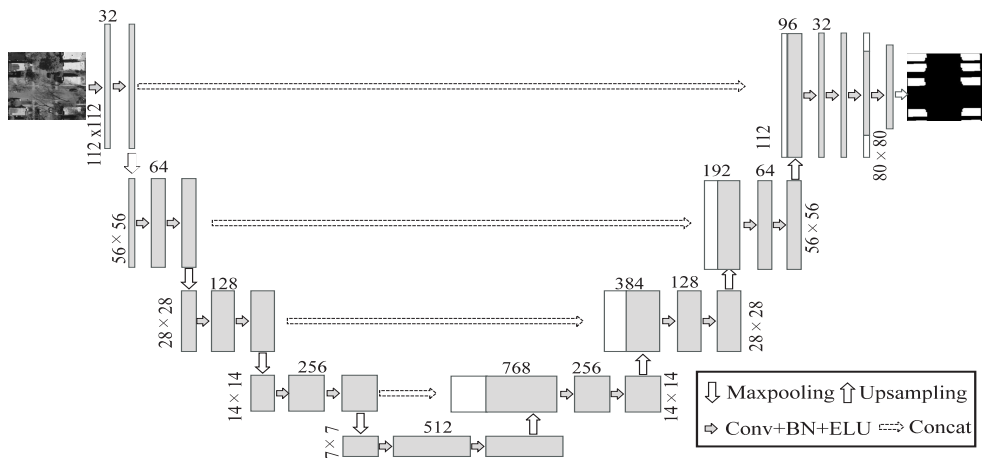


图2 U-Net 分割网络结构

2.3 损失函数及模型融合

上述两个模型的损失函数采用与 Sigmoid 函数对应的 Binary Cross Entropy 函数,并通过负对数,将输出的概率值从最大化概率变成最小化信息熵。其表达式如下:

$$H = -\frac{1}{m} \sum_{i=1}^m [y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad (2)$$

其中, y 表示像素值的真实值; \hat{y} 表示像素值的预测值。

此外,为了使建筑物分割任务有更高的精度,在二进制交叉熵损失函数中加入 IOU 公式,形成联合损失函数。最终两个模型的损失函数如下:

$$L = H - \log J_m \quad (3)$$

$$J_m(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^m \frac{y_i \cdot \hat{y}_i}{y_i + \hat{y}_i - y_i \cdot \hat{y}_i} \quad (4)$$

受模型融合方法的启发,将两种模型的预测结果通过加权的方式融合,即保留编解码分割网络对大建筑物的分割提取,也保留了 U-Net 分割网络在小型建筑物上的精确分割。结合式1,模型预测融合的公式如下:

$$P = \alpha * p_e + \beta * p_u, \alpha + \beta = 1 \quad (5)$$

其中, p_e 为编解码网络的预测结果; p_u 为 U-Net 网络的预测结果。

模型融合流程如图3所示。在输入的影像中以给定点为中心,分别提取尺度大小为 112^2px 和 128^2px 的影像块,分别送入两个模型中进行预测,经过模型运算后得到大小为 80^2px 的预测结果,最后对预测结果按照上述公式进行加权叠加得到最终的分割掩码图像。

并且方差为1,加速网络收敛速度。紧接着采用 ELU 非线性激活函数。下采样操作采用 max-pooling 来降低特征图分辨率。在上采样阶段,使用 Concat 连接层将包含上下文信息的低层级的特征图和包含图像细节的高层级的特征图结合起来。最后将剪裁得到的特征输入到 Sigmoid 层,通过反向传播进行训练。

实验表明,当 $\alpha = 0.45$, $\beta = 0.55$ 时,融合模型能取得最好结果。

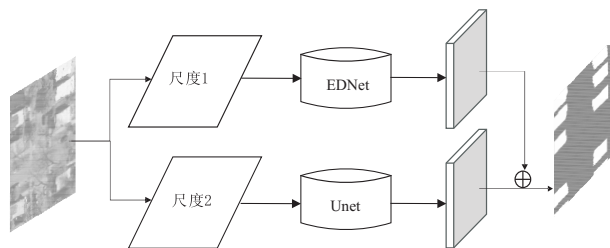


图3 模型融合流程

3 实验

3.1 实验数据

本次实验的数据集^[15]来自法国国家信息与自动化研究所(Inria),该数据集包含来自美国和澳大利亚不同城市居民区上空的遥感影像,它们被标注为建筑物和非建筑物两类。与一般数据集不同的是,该数据集的训练样本和测试样本按影像所属地区划分,而不是将同一地区的不同区域分别划分为训练样本和测试样本。也就是说,用来自 Chicago 的训练数据得到的分类器可以用来预测 San Francisco 的数据,旨在提高模型的泛化能力。该数据集的航空正射校正影像的空间分辨率是0.3 m,具体信息如图4所示。

为了评价在该数据集上建筑物的分割精度,本次实验定义了两种评价方法,分别是像素精度(accuracy)和 IOU (intersection over union)^[16]。假设图像数据集有 n_c 个语义类别的物体, n_{ij} 表示真实语义类别为 i 中被预测为类别 j 的像素数,用 $t_i = \sum_j n_{ij}$ 表示

类别 i 的像素总数(正确预测的像素数与错误预测为

其他语义类别的像素数的总和),则有:

Train	Tiles	Total area
Austin, TX	36	81 km ²
Chicago, IL	36	81 km ²
Kitsap County, WA	36	81 km ²
Vienna, Austria	36	81 km ²
West Tyrol, Austria	36	81 km ²
Total	180	405 km ²

Test	Tiles	Total area
Bellingham, WA	36	81 km ²
San Francisco, CA	36	81 km ²
Bloomington, IN	36	81 km ²
Innsbruck, Austria	36	81 km ²
East Tyrol, Austria	36	81 km ²
Total	180	405 km ²

图 4 数据集样本统计(每张图大小为 5 000²px)

$$\text{accuracy} = \frac{\sum_i n_{ii}}{\sum_i t_i} \tag{6}$$

$$\text{IOU} = \frac{1}{n_c} \sum_i \frac{n_{ij}}{t_i + \sum_j n_{ij} - n_{ii}} \tag{7}$$

3.2 实验设计

对于所提出的两种模型,使用 Inria 数据集中的 180 幅高分辨率影像及其对应的语义标注图进行学习。首先,从 180 幅影像中随机切割出 144 块大小为 128×128 像素的图像块,并随机进行图像的水平、竖直或者镜像翻转,组成维度为 144×128×128 的图像斑块作为编解码分割网络的输入。在一次迭代的过程中,使用 5 000 批上述维度图像斑块进行网络参数的批量梯度下降学习,共进行 25 轮迭代。对于 U-Net 分割网络模型,在 25 轮迭代过程中,使用维度为 64×128×128 的图像斑块作为模型输入,每轮迭代使用 8 000 批图像斑块进行训练。在两个模型中,使用 Nadam 优化器以初始学习率 0.001 进行网络参数学习。完成学习后,使用 test time augmentation 对测试集中 180 幅影像进行预测,每张影像分别进行水平旋转预测、竖直翻转预测、镜像翻转预测和直接预测,将预测得到的四幅图像复原后,取各图像像素值的平均值,并以 0.5 的预测阈值得到最终的二分类分割图像。最后,将得到预测的语义分割图与对应 Ground Truth 进行比对,分别得到每个地区的建筑物分割精度和 IOU。

3.3 实验结果

本次实验是基于 Keras 深度神经网络框架,使用单张 12 GB GTX TitanX 显卡完成模型的训练。模型训练结束后,给出模型在测试集上三个不同区域上的样例结果。两个模型预测结果和模型融合的预测结果如图 5 所示。图中从上到下分别是 RGB 遥感影像、Ground Truth、编解码网络、U-Net 网络和模型融合的结果。可以看出,编解码网络能更好地识别和提取大型建筑物的轮廓,而 U-Net 网络由于结合了高低层级特征图,能更好地提取建筑物的边缘细节和小型建筑物轮廓。模型融合很好地结合了二者优点。同时在结果上出现了小面积的 false positive(将非建筑物识别为

建筑物)和 false negative(无法识别建筑物区域)。

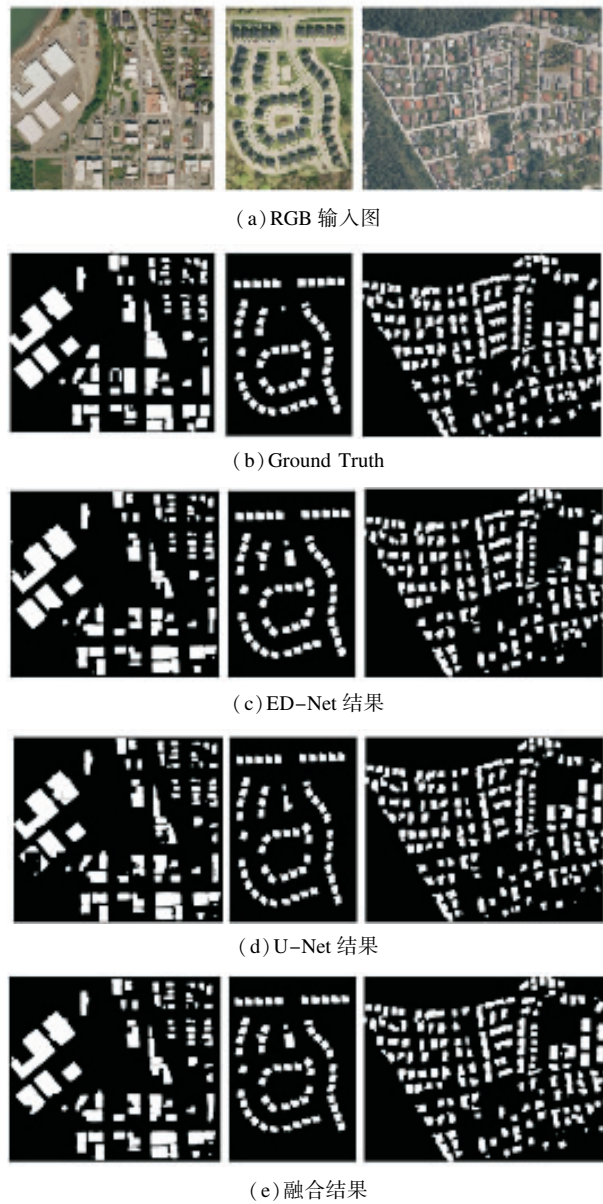


图 5 模型在测试集上的预测结果

表 1 给出了三个模型在测试集上的量化结果,按照区域显示预测精度和 IOU,并求取区域预测的结果的平均值作为最后的预测结果。从表中可以看出,模型融合在测试集上的整体结果相对 U-Net 有了近似 2% 的提升,达到了 70.08%,也就是说预测区域的

面积覆盖了真实区域面积的 70%。对于特定地区,例如 Bloomington, IOU 指数有 3.46% 的提升。

表 1 模型在测试集上的定量评估结果

指标		Bellingham	Bloomington	Innsbruck	San Francisco	East Tyrol	Overall
ED-Net	IoU	68.97	49.60	66.87	58.30	73.45	61.91
	Acc.	96.95	95.44	96.21	86.66	97.68	94.59
U-Net	IoU	69.94	57.57	69.92	68.71	74.55	68.14
	Acc.	96.92	96.06	96.43	89.63	97.70	95.35
模型融合	IoU	71.59	61.03	70.39	70.41	76.97	70.08
	Acc.	97.14	96.33	96.53	90.21	97.94	95.63

4 结束语

文中设计和实现了两种基于高分辨率遥感影像的建筑物语义分割模型,进行了模型融合,并通过实验对模型进行了验证和分析。实验结果表明,由于输入编解码分割网络的图像斑块尺度较大,包含更多的上下文细节,使得编解码网络能更好地提取大型建筑物的轮廓。而 U-Net 网络,结合了高层级的图像细节特征,对于小型建筑的特征更加敏感。模型融合结合了二者优点,在建筑物分割测试集上可以获得更加令人满意的语义分割结果。

参考文献:

[1] 孙家柄. 遥感原理与应用[M]. 第 2 版. 武汉:武汉大学出版社,2009.

[2] 巫兆聪,胡忠文,张 谦,等. 结合光谱、纹理与形状结构信息的遥感影像分割方法[J]. 测绘学报,2013,42(1):44-50.

[3] 莫 玉. 高光谱图像特征学习与分类算法研究[D]. 西安:西安电子科技大学,2015.

[4] 王巧玉. 基于深度学习的高光谱遥感图像分类[D]. 泉州:华侨大学,2016.

[5] 魏德强. 高分辨率遥感影像建筑物提取技术研究[D]. 郑州:解放军信息工程大学,2013.

[6] 付 卓,胡吉平,谭衢霖,等. 遥感应用分析中影像分割方法[J]. 遥感技术与应用,2006,21(5):456-462.

[7] 代具亭,汤心溢,刘 鹏. 基于深度学习的语义分割网络[J]. 红外,2018,39(4):33-38.

[8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2015:3431-3440.

[9] ZHENG S, JAYASUMANA S, ROMERA-PAREDES B, et al. Conditional random fields as recurrent neural networks[C]//IEEE international conference on computer vision. Santiago, Chile: IEEE, 2016:1529-1537.

[10] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Seg-Net: a deep convolutional encoder-decoder architecture for scene segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12):2481-2495.

[11] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer, 2015:234-241.

[12] 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.

[13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2016:770-778.

[14] WANG S, FIDLER S, URTASUN R. Holistic 3D scene understanding from a single geo-tagged image[C]//Computer vision and pattern recognition. Boston, MA, USA: IEEE, 2015:3964-3972.

[15] MAGGIORI E, TARABALKA Y, CHARPIAT G, et al. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark[C]//IEEE international geoscience and remote sensing symposium. [s. l.]: IEEE, 2017:3226-3229.

[16] 刘 丹,刘学军,王美珍. 一种多尺度 CNN 的图像语义分割算法[J]. 遥感信息, 2017, 32(1):57-64.