

文章编号: 2096-1618(2022)01-0046-05

一种高分辨率卫星图像道路提取方法

晏美娟, 魏 敏, 文 武

(成都信息工程大学计算机学院, 四川 成都 610225)

摘要:为实现高分辨率卫星图像的道路自动提取,设计一种编码器-解码器结构的图像分割方法。针对卫星图像中乡村地区的道路提取结果不佳,以及不能对阴影区域、被遮挡区域的道路进行有效提取的问题,以不含全连接层的 VGG13 作为编码器的骨干网络,对解码器部分进行设计,达到对道路区域进行有效提取的目的,并对模型训练使用的损失函数进行介绍。在开始训练之前,对 DeepGlobe 道路提取数据集进行预处理。使用 PaddlePaddle 深度学习框架展开实验,改进后的方法在验证集上的 IoU, acc, Kappa 分别可以达到 0.6194, 0.9811, 0.7551, 对比实验结果显示,与使用 DeepLabv3+ 和 U-Net 的道路提取方法相比,可以有效提升道路提取结果的准确性和完整性。

关 键 词:深度学习;卫星图像;道路提取;图像分割;飞桨

中图分类号:TP75

文献标志码:A

doi:10.16836/j.cnki.jcuit.2022.01.008

0 引言

在处理广泛的计算机视觉问题时,卷积神经网络都可以作为有用的模型^[1]。高分辨率的卫星图像为线性特征的提取提供了新的可能性,道路就是一个例子^[2]。作为遥感领域的一个基础任务,近年来,道路提取已经成了一个研究的热门话题,尽管得到了广泛研究,但由于道路结构的特殊性,从高分辨率的卫星图像中精确提取道路特征仍然是一个具有挑战性的任务^[3]。在卫星图像中,道路通常是细长的复杂结构,在整张图片中占比很小,因此保存好详细的空间特征是意义重大的。

在计算机视觉领域中,对图像中的感兴趣区域进行有效分割,即对图像中的每一个像素点进行精确的类别划分,是一个具有挑战性的研究任务,而针对卫星图像的分割难度更大。对高分辨率的卫星图像进行分析已经成为当下的一个研究热点,而从卫星图像中提取道路具有重要的研究价值,可以应用于城市规划、路线图更新、车辆导航、地理信息更新等领域。最直接的提取方法就是人工分割,但这种方法费时费力。现在,很多计算机视觉相关的研究任务都在尝试利用深度学习技术来对传统方法进行改进^[4]。

1 研究现状

通过一代代航天科技工作者的不懈努力,卫星发

射技术飞速发展,卫星上安装的设备也越来越先进,将道路特征提取的深度学习方法运用于高分辨率的卫星图像中越来越具有高效性和经济性^[5]。在卫星图像道路提取领域,研究者提出了各种方法,主要可以归结为 3 类:(1)像素级的分割方法^[6-8];(2)道路中心线提取^[9-10];(3)前两类方法的结合^[11-12]。

Volodymyr Mnih 等^[13]将限制玻尔兹曼机用于分割高分辨率的航空图像中的道路区域,通过使用无监督学习方法及利用输出标签的局部空间一致性来初始化特征检测器,引入后处理程序用于显著提升神经网络的预测结果,但模型不能处理大型建筑造成的阴影和遮挡。Tao Sun 等^[14]连接两个 U-Net,实现多输出,第一个 U-Net 输出辅助信息,如道路拓扑图,第二个 U-Net 产生道路掩码,另外使用了后处理方法,包含路线图向量化和具有层次阈值的最短路径搜索,可以连接损坏的道路区域,但最大的问题就是乡村地区的预测效果不佳。

本文使用 PaddlePaddle 深度学习框架进行实验,基于 U-Net 网络结构对道路提取的方法进行研究,使用 VGG13 作为编码器,并对解码器结构做相应的改进。将 dice_loss 和 bce_loss 作为损失函数。将道路提取任务视为一个二分类的图像分割任务,从而生成像素级的道路预测结果图,对于乡村地区比较狭窄的泥路也有较好的识别性,对阴影区域的道路及被树木、房屋遮挡的道路进行提取时,本文方法也具有鲁棒性。

2 研究方法

受自编码器的启发,现有的很多语义分割技术都

使用编码器-解码器结构作为其网络架构的核心^[15]。道路识别任务就是一种语义分割任务,研究者基于此方向进行了大量的实验。Oleksandr Filin 等^[16]认为编码器-解码器结构的神经网络更可取,其中最具代表性的是 U-Net 和 SegNet。更深的网络层通常包含高级别的特征,浅层则携带更多细节和位置信息。U-Net 可以使用非常少的图片进行端到端的训练^[17],利用跳跃连接将低层特征和高层特征相融合,使解码器可以利用浅层网络层的信息,实现精确的像素级别的定位功能,同时,也有利于训练的收敛,因为梯度可以直接向网络浅层传递。

经典 U-Net 模型结构简单,且具有良好的分割效果,故本文基于 U-Net 网络结构进行改进,编码器部分使用 VGG13 作为骨干网络,使网络对道路特征进行提取的能力有所提高。用卷积层代替全连接层,将卷积神经网络视为一个强有力的特征提取器,使网络输出空间特征图,而不是输出分类得分。编码器部分两次卷积操作和一次最大池化操作交替进行,第一次卷积操作之后增大特征图的通道数,池化操作之后特征图的尺寸减小,去除冗余信息,实现对特征的压缩。在解码器部分,上采样操作之后进行两次卷积操作。将上采样之后的特征与编码器部分高分辨率的特征通过跳跃连接相结合,以实现精确的定位功能。解码器部分上采样方式为双线性插值,目的是逐步增大特征图的分辨率,使其最后与原图大小一致。

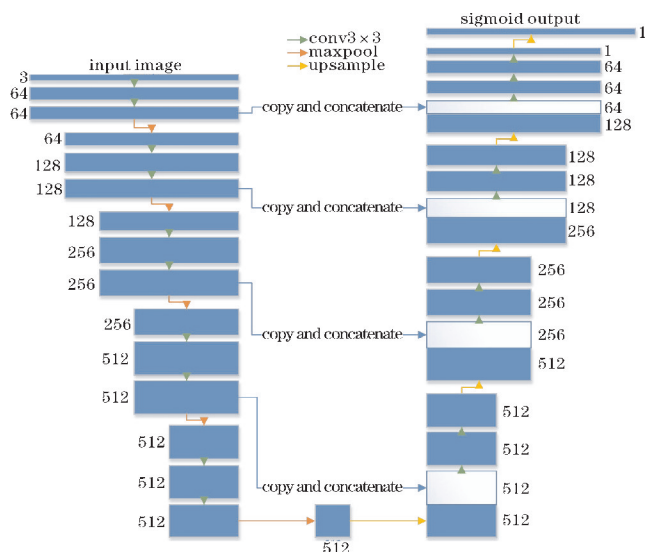


图 1 本文网络结构

如图 1 所示,使用不包含全连接层的 VGG13 网络作为本文网络结构的编码器部分的骨干网络,一条矩形块代表着一个特征图,特征图的尺寸与矩形块的宽度相关,该特征图拥有多通道数,矩形块的高度则与通道数相关,数字为具体的通道数信息。与编码器部分

相反,解码器部分特征图的通道数逐渐减小。网络中间部分的箭头代表从指定编码层连接到相应解码层的信息传输。在卷积操作之后、激活函数之前使用 batch normalization (BN)。

以一个通道数为 512 的特征图作为网络的中间部分,从视觉上有效分隔网络的编码器部分和解码器部分。上采样操作之后特征图的通道数维持不变,且上采样之后的特征图将与对应编码器部分的特征图相融合。与池化操作对应,网络包含 5 次上采样操作,使输出的提取结果图尺寸和输入图像的尺寸一致。

Karen Simonyan 等^[18]将 3×3 的卷积核贯穿整个网络,使 VGG 系列网络需要的参数量更少。简单来说,采用两层 3×3 的卷积核可以使特征图的感受野大小为 5,而比直接采用 5×5 的卷积核需要的参数量更少。VGG13 网络包含 10 层 3×3 的卷积核,外加 3 层全连接层,每两个卷积层之后进行一次最大池化操作,此操作之后特征图的宽度和高度都会减半,而通道数则会成倍增加,即由 64 增至 128,再增至 256,最后增加至 512,并维持不变。

本文使用 dice coefficient loss (dice_loss) 和 binary cross entropy loss (bce_loss) 作为损失函数。在二分类图像分割任务中,经常出现类别分布不均的情况,即此类分割任务一般背景所占的比例远大于前景所占的比例,而 dice_loss 通过预测像素和标注像素的交集除以它们的总体像素进行计算,将一个类别的全部像素作为一个整体进行考量,因而不受大量背景像素的影响,如式(1)所示。在实际应用中,dice_loss 通常与 bce_loss 组合使用,从而提高模型训练的稳定性。

$$\text{dice_loss} = 1 - \frac{2|GT \cap P|}{|GT| + |P|} \quad (1)$$

其中,GT 表示输入图像的 ground truth, P 代表网络输出的预测结果。 $\|\cdot\|$ 表示矩阵元素之和。 $|GT \cap P|$ 表示 GT 和 P 共有的元素数目,在实际运算时,通过求两者的逐像素乘积之和进行计算。

3 实验与结果分析

3.1 评价指标

在图像分割领域中,可以通过准确率(acc)、平均交并比(mean intersection over union, mIoU) 和 Kappa 系数这 3 个指标对模型质量进行评估。acc 为类别(在本文中有道路类和背景类)预测正确的像素数目占图像总体像素的比例。求 mIoU 时对每个类别单独进行推理计算,如式(2)所示,然后取所有类别结果的

平均值,可以比较准确地反映模型总体的预测性能。

$$IoU = \frac{P \cap GT}{P \cup GT} \tag{2}$$

其中, P 表示类别的预测区域, GT 表示类别的实际区域。

在实践中通常表示为

$$IoU = \frac{TP}{TP+FP+FN} \tag{3}$$

其中, TP 即 true positive, FP 即 false positive, FN 即 false negative。

Kappa 系数基于混淆矩阵进行运算,如式(4)所示,其取值在 $-1 \sim 1$,通常情况下大于 0。

$$Kappa = \frac{P_c - P_r}{1 - P_r} \tag{4}$$

其中 P_c 代表分类器的准确率, P_r 代表随机分类器的准确率。Kappa 系数的值越高意味着模型的质量越好。

3.2 实验设计

实验使用了 6226 张 3 通道的卫星图像,即图像格式为 RGB,这些图像来自 DeepGlobe 道路提取数据集,每张图片为 1024×1024 像素,地面分辨率为每像素对应 $50 \text{ cm} \times 50 \text{ cm}^{[19]}$,每张图像对应着一张黑白的标注图,黑色代表背景,白色代表道路。本文将卫星图像和其相应的标注图像按比例进行子集的划分,使训练集包含 4360 张卫星图像,验证集和测试集各包含 933 张卫星图像。为防止训练数据过拟合^[20],本文采用镜像的数据增强方式来适当增加训练集的大小。训练时将输入图像大小设置为 512×512 像素,验证时亦然。

实验电脑的 CPU 为 AMD Ryzen 5 3600X 6-Core Processor;内存为 16.0 GB ;GPU 为 NVIDIA GeForce RTX 2070 SUPER,包含 8.0 GB 专用 GPU 内存。采用 Adam^[21] 优化器,设置初始学习率的值为 0.001。实验采用 Jupyter Notebook 集成开发环境,选择 PaddlePaddle 深度学习框架。由于在此框架中 255 为忽略的像素标签值,故对原始标注图进行预处理,使背景的像素值为 0,前景的像素值为 1,此时道路显示为红色,背景为黑色。

使用不同的 BATCH_SIZE 对本文网络进行训练,结果如表 1 所示。

表 1 不同 BATCH_SIZE 下本文方法的评价指标对比

BATCH_SIZE	IoU	acc	Kappa
1	0.5195	0.9747	0.6707
2	0.6031	0.9805	0.7423
4	0.6194	0.9811	0.7551

在道路提取任务中,分类类别仅有两类,即道路类和非道路类,故在文中使用道路类的 IoU 作为模型的评价指标,根据表 1 的实验结果,在对网络进行训练时,将一次选取的样本数(BATCH_SIZE)设置为 4。

3.3 结果与分析

如图 2 所示,第一列为输入网络的卫星图像,第二列为经过处理的相应的标注图,第三列为使用本文网络提取的道路结果图,第四列为使用 U-Net 模型提取的道路结果图,最后一列为使用 DeepLabv3+模型提取的道路结果图,网络输出的预测结果图尺寸和输入图像的尺寸一致。

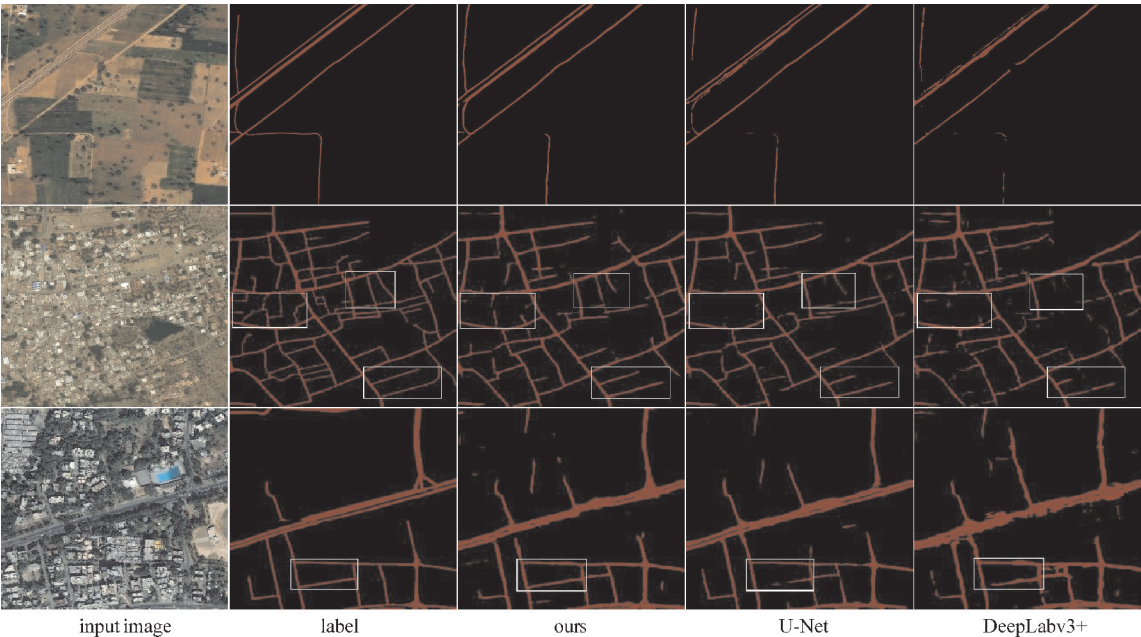


图 2 3 种道路提取方法的预测结果对比

可以看出,经过 100 轮的训练之后,3 种模型的预测质量有所区别,使用本文方法进行道路提取,能够有效提升道路分割结果的精确度和完整度。针对乡村地区,如第一行图像所示,本文方法可以有效提取树木遮挡的道路区域,对于比较狭窄的泥路也有较好的识别性;在道路网络复杂的城镇区域,如第二行白色方框选中的区域所示,对树木、房屋遮挡的道路进行提取时,本文方法也具有鲁棒性。在有大型建筑、高大树木的区域,如第三行白色方框选中的区域所示,对于建筑、树木造成的阴影区域的道路也可以进行有效的提取。

表 2 3 种道路提取方法的评价指标对比

Model	IoU	acc	Kappa
DeepLabv3+	0.5550	0.9774	0.7021
U-Net	0.5982	0.9811	0.7389
ours	0.6194	0.9811	0.7551

如表 2 所示,本文道路提取方法与另外两种方法相比,具有显著的优势。使用本文方法进行道路提取时,IoU 值比使用 U-Net 模型提升了 0.0212,相比使用 DeepLabv3+模型提升了 0.0644;Kappa 值比使用 U-Net 模型提升了 0.0162,相比使用 DeepLabv3+模型提升了 0.0530;acc 值和使用 U-Net 模型时相同,相比使用 DeepLabv3+模型提升了 0.0037。实验证明,本文方法可以有效提高道路提取的质量,且网络的训练耗时大大减少,使用 U-Net 方法进行训练时,耗时 14h 28m,而本文网络训练仅耗时 8h 16m。

4 结束语

本文使用具有跳跃连接的编码器-解码器结构对卫星图像中的道路进行自动提取,编码器部分将无全连接层的 VGG13 作为骨干网络,并对解码器部分进行设计。网络使用 dice_loss 和 bce_loss 作为损失函数。未来可尝试使用其他骨干网络,进一步提高道路提取的精度、IoU 和速度。Volodymyr Mnih 等证明无监督预训练方法和监督后处理方法可以在很大程度上提升道路检测器的性能,未来也可以对相关方法进行研究。针对树木遮挡导致的特征信息缺失问题,将道路的几何特征信息作为补充以确保道路提取的完整性也是未来的一个发展趋势^[22]。在未来的研究中,也可以在道路区域提取的基础上增加对道路中心线的提取。

参考文献:

[1] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation

Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(8):2011–2023.

[2] Buslaev A V, Seferbekov S S, Iglovikov V I, et al. Fully Convolutional Network for Automatic Road Extraction from Satellite Imagery[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018.

[3] Yuewu H, Zhaoying L, Ting Z, et al. C-UNet: Complement UNet for Remote Sensing Road Extraction[J]. Sensors,2021,21(6):2153.

[4] Iglovikov V, Shvets A. TerausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation [DB/OL]. <https://arxiv.org/pdf/1801.05746.pdf>,2018–01–17.

[5] Pan D, Zhang M, Zhang B. A Generic FCN-based Approach for the Road-Network Extraction from VHR Remote Sensing Images-Using OpenStreet-Map as Benchmarks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2021,14:2662–2673.

[6] Zhang Z, Liu Q, Wang Y. Road Extraction by Deep Residual U-Net[J]. IEEE Geoscience and Remote Sensing Letters,2018,15(5):745–753.

[7] 李代栋,赫晓慧,李盼乐,等. 基于 SPUD-ResNet 的遥感影像道路提取网络[OL]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20201111.1410.010.html>. 计算机工程与应用,2020:1–10.

[8] Chen Z, Wang C, Li J, et al. Reconstruction Bias U-Net for Road Extraction from Optical Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2021,14:2284–2294.

[9] Bastani F, He S, Abbar S, et al. RoadTracer: Automatic Extraction of Road Networks from Aerial Images[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018.

[10] 蒋星详,肖莉. 一种多特征融合的高分辨率遥感影像道路中心线提取算法[J]. 测绘地理信息,2019,44(4):98–101.

[11] Shao Z, Zhou Z, Huang X, et al. MRENet: Simultaneous Extraction of Road Surface and Road Centerline in Complex Urban Scenes from Very High-Resolution Images [J]. Remote Sensing, 2021,13(2).

[12] Mattyus G, Luo W, Urtasun R. DeepRoadMapper:

- Extracting Road Topology from Aerial Images [C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017.
- [13] Mnih V, Hinton G E. Learning to detect roads in high-resolution aerial images [C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010.
- [14] Tao S, Chen Z, Yang W, et al. Stacked U-Nets with Multi-output for Road Extraction [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018.
- [15] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation [DB/OL]. <https://arxiv.org/abs/1707.03718>, 2017-06-14.
- [16] Filin O, Zapara A, Panchenko S. Road Detection with EOSResUNet and Post Vectorizing Algorithm [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018.
- [17] 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, 2015.
- [18] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [DB/OL]. <https://arxiv.org/pdf/1409.1556.pdf>, 2015-04-10.
- [19] Demir I, Koperski K, Lindenbaum D, et al. DeepGlobe 2018: A Challenge to Parse the Earth-through Satellite Images [DB/OL]. <https://arxiv.org/pdf/1805.06561.pdf>, 2018-05-17.
- [20] Zhou L, Zhang C, Ming W. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018.
- [21] Kingma D, Ba J. Adam: A Method for Stochastic Optimization [DB/OL]. <https://arxiv.org/abs/1412.6980>, 2017-01-30.
- [22] Wu Q, Luo F, Wu P, et al. Automatic Road Extraction from High-Resolution Remote Sensing Images using a Method Based on Densely Connected Spatial Feature-Enhanced Pyramid [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 14: 3-17.

A Method of Road Extraction for High-Resolution Satellite Images

YAN Meijuan, WEI Min, WEN Wu

(College of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In order to achieve the automatic road extraction of high-resolution satellite images, an image segmentation method based on the encoder-decoder structure is designed. In view of the poor results of road extraction in rural areas in satellite images, and the inability to effectively extract roads in shadow areas and sheltered areas, VGG13 without full connection layers is used as the backbone network of the encoder, the decoder is designed to effectively extract road areas, and the loss functions used in the training of model are introduced. Before starting the process of training, the DeepGlobe road extraction data set is preprocessed. The PaddlePaddle deep learning framework is used to carry out experiments. On the validation set, the value of IoU, acc and Kappa of the improved method can reach 0.6194, 0.9811 and 0.7551 respectively. The experimental results show that this method can effectively improve the accuracy and integrity of road extraction results compared with the road extraction methods using DeepLabv3+ and U-Net.

Keywords: deep learning; satellite imagery; road extraction; image segmentation; PaddlePaddle