

doi: 10. 6046/gtzyyg. 2020162

引用格式: 刘钊,赵桐,廖斐凡,等. 基于语义分割网络的高分遥感影像城市建成区提取方法与对比分析[J]. 国土资源遥感,2021,33(1):45-53. (Liu Z,Zhao T,Liao F F,et al. Research and comparative analysis of urban built - up area extraction methods from high - resolution remote sensing images based on semantic segmentation network [J]. Remote Sensing for Land and Resources , 2021,33(1):45-53.)

基于语义分割网络的高分遥感影像城市建成区提取方法与对比分析

刘 钊, 赵 桐, 廖斐凡, 李 帅, 李海洋
(清华大学土木工程系交通工程与地球空间信息研究所,北京 100084)

摘要:城市建成区的提取对城市发展规划有着重要的作用。为了找出能兼顾效率和识别准确率的基于卷积神经网络的遥感影像城市建成区提取方法,从神经网络结构的原理出发,对多种语义分割网络的内部结构进行对比分析,并针对语义分割网络分别进行训练及结果比较。实验结果表明,ShelfNet-50 网络能够在训练速度最快的同时保证很高的识别准确率,在训练时长仅需 14 h 的同时达到了 77% 的前景分割精度,且 ShelfNet-50 网络预测的结果也与相应的遥感影像数据高度吻合。实验说明 ShelfNet-50 网络可应用于高分遥感影像的城市建成区提取研究。

关键词:高分遥感影像;卷积神经网络;语义分割;城市建成区

中图法分类号:P 237 **文献标志码:**A **文章编号:**1001-070X(2021)01-0045-09

0 引言

近年来,中国的城市化进程突飞猛进,而现实城市发展进程的指标之一就是城市建成区。根据国家质量技术监督局和国家建设部共同发布的《城市规划基本术语标准》,建成区定义为:城市行政区内实际已成片开发建设、市政公用设施和公共设施基本具备的区域^[1]。城市建成区是反映城市综合经济实力和城市化水平的重要指标。获取历年以来城市建成区数据,对于城市的建设和管理企业的决策等具有重要的意义。

城市建成区的提取是近年来遥感应应用领域研究的热点内容。传统的研究方法主要包括都市化指数(urban index, UI)^[2]、归一化建筑指数(normalized difference built-up index, NDBI)^[3]、归一化植被指数(normalized difference vegetation index, NDVI)^[4]等方法。这些指数不仅可以单独运用,也可相互结合综合运用^[5],能够得到更好的效果。传统方法由于是通过地物的光谱特征提取关键信息,经常出现同谱异物或者同物异谱的现象,从而使该类方法很不稳定,在识别地物时容易存在较大误差。

机器学习方法中,随机森林法^[6]、支持向量机法^[7]以及无监督的聚类方法^[8]同样可以完成建成区的识别。Gong 等^[9]应用哨兵 2 号影像采用随机森林算法,在 Google Earth Engine 平台上达到了 72.6% 的高精度。机器学习的算法优点在于算法简单,运行速度快。但算法简单带来的缺陷就是层次太浅,无法处理复杂的分类问题,以及面对大数据量的问题时,预测的精度会受到限制。所以,传统机器学习算法并不是当今热点问题的最优解。

深度学习算法很好地解决了传统机器学习算法中层次浅的问题,成为了当今的研究热点。卷积神经网络^[10]则是深度学习中应用最广泛的分支,目前已被大量用在城市建成区的识别之中。现代意义上的卷积神经网络起源于 2012 年的 AlexNet^[11]网络,该网络相比于以往的神经网络拥有更深的层次和新的激活函数 ReLU,从而加快了训练速度。卷积神经网络在图像分类上取得了高精度后,随即被应用到其他问题,例如图像语义分割以及图像实例分割。其中图像语义分割类似于图像分类,不同点在于语义分割是针对每一个像素的分割问题。

近几年的研究和应用中涌现出了很多优秀的语义分割网络,每年图像分割竞赛的语义分割网络精

收稿日期: 2020-06-09; 修订日期: 2020-12-01

第一作者: 刘 钊(1967-),男,副教授,主要从事 GIS 及其应用、云 GIS、时空大数据及遥感图像处理等方面的研究。Email: liuz@tsinghua.edu.cn。

通信作者: 赵 桐(1996-),男,硕士研究生,主要从事 GIS 和深度学习研究。Email: zhaot18@mails.tsinghua.edu.cn。

度也在逐年提升。很多国内外学者应用语义分割网络进行遥感影像建成区的识别,并取得了理想的结果。杨建宇等^[12]运用 SegNet^[13] 语义分割网络在 WorldView 影像上进行霸州地区农村用地的提取,并得到了很高的分类精度;苏健民等^[14]将 U-Net^[15] 语义分割网络应用在 CCF 卫星影像数据集上,取得了 90% 的测试准确率;Chen 等^[16]基于空洞卷积^[17]的理论提出了 Deeplab v3 模型结构,将准确率进一步提升;王俊强等^[18]也利用 Deeplab v3 网络对道路和建筑物进行提取,得到了较高的分割精度。

在一个典型的语义分割网络之中,每一个卷积层之间的连接关系会影响到最终训练的精度。残差网络(residual neural network, ResNet)^[19]通过跳接的方式使网络的训练精度更高,解决了传统的 AlexNet 和 VGGNet 网络最大的缺陷。本文以 ResNet50 以及 ResNet101 残差网络作为解码器的基础,对 Deeplab v3,金字塔场景解析网络(pyramid scene parsing network, PSPNet)^[20]和 ShelfNet^[21]3 个以 ResNet 网络为骨架的语义分割网络的内部网络结构进行对比分析,并通过同一高分遥感影像城市建成区数据进行对比实验,比较得出提取城市建成区时最适用的语义分割网络。

1 ResNet 网络原理分析

ResNet 由 He 等提出^[19]。该网络在 ILSVRC 比赛中取得冠军,分割效果非常理想。与早期的 AlexNet 和 VGGNet^[22]相比,能够在保持更低参数数量的同时得到更好的效果,并且解决了 VGGNet 及其他更早的卷积神经网络中的诸多缺陷。AlexNet 网络和 VGGNet 网络仅仅是由上而下的编码结构,此类结构的层数通常会存在最大值,而超过最大值的网络很容易出现过拟合的问题,从而导致精度的降低。ResNet 最大的创新点在于网络内部的残差单元,其基本结构如图 1 所示。残差单元保证了 ResNet 网络可以达到很深的层次。图中的 x 为残差学习模块的输入层, $F(x)$ 为残差学习网络的映射。

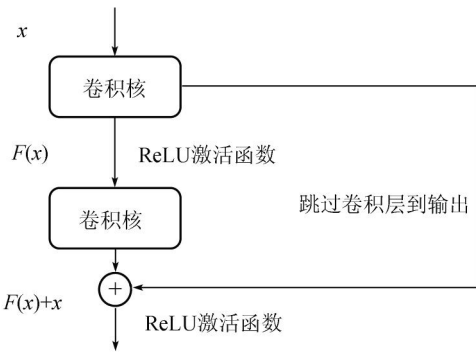


图 1 ResNet 基本结构
Fig. 1 ResNet basic structure

传统的 AlexNet 和 VGGNet 网络最大的缺陷在于,当网络变得越来越深的时候,训练误差曲线会趋于平缓,准确率不再上升,VGGNet 达到 19 层甚至会导致分类性能的下降。ResNet 很好地解决了这类问题:它不再用堆叠的卷积层直接拟合期望的特征映射,而是将原本的映射和输入本身做差,拟合残差映射。假设卷积层的映射为 $H(x)$,那么残差映射 $F(x) = H(x) - x$ 。ResNet 网络将输入的信息绕道直接传到输出端,这样做的优势在于浅层的信息成功进入了深层的卷积层,使得整个卷积网络中融合了大量浅层信息,避免了梯度消失问题。ResNet 网络常用层数有 34 层、50 层、101 层以及 152 层。

由于 ResNet 网络的优异表现,现如今很多语义分割网络都基于 ResNet 网络搭建,包括 Deeplab v3 网络、PSPNet 网络以及 ShelfNet 网络等。它们都将 ResNet 作为基本骨架,再加上自身的创新点,从而形成了各自网络的优点。

2 语义分割网络原理及对比分析

2.1 Deeplab v3 网络原理

Deeplab v3 语义分割网络中设计了不同采样率的空洞卷积并将其进行融合。架构中的空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块可以提取不同尺度上特征图的卷积特征,从而使准确率得到进一步的改善。Deeplab v3 语义分割网络的原理示意图如图 2 所示。

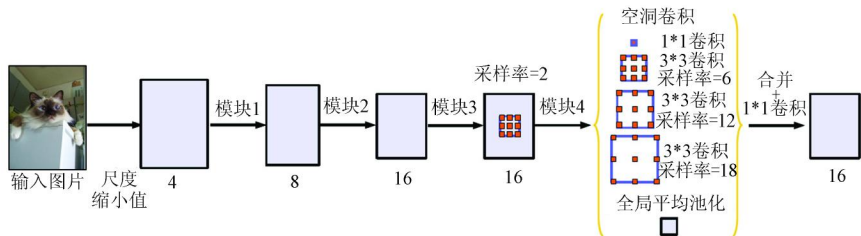


图 2 Deeplab v3 语义分割网络结构^[17]

Fig. 2 Deeplab v3 semantic segmentation network structure^[17]

Deeplab v3 网络主要解决了 2 个传统语义分割网络的缺陷：一是不能很好地顾及不同尺度物体的信息,很容易出现漏识别小物体或误识别大物体的情形；二是当网络本身层次太深,卷积核本身又不太大的情况下,很容易出现漏识别,以及识别结果分辨率过低的情况。

在连续下采样的过程中,图片越小,细节信息丢失越多,这对于语义分割是不利的。故在 ResNet 残差网络得到输出后,Deeplab v3 网络在卷积的最后一层直接通过某一采样率的空洞卷积保持输出步幅(图 2 中采样率 = 2)。这样既保证了图片质量,也没有增加参数。ASPP 模块的设计则更为高级。ASPP 金字塔池化模块通过 3 个不同采样率的 3 * 3 空洞卷积和一个 1 * 1 卷积保证不同尺度物体的信息。同时,在采样率接近特征图的大小时,3 * 3 的滤波器不能很好发挥捕捉全图的作用,故 Deeplab

v3 网络在最后加入全局平均池化,输出 256 个通道,经过卷积以后再和其他特征图融合起来。该方法很好地提升了分类效果。

然而,Deeplab v3 网络同样存在缺陷,即在编码过程结束后直接解码到原尺寸,这可能会导致输出结果放大的效果不好,信息太少。其次,Deeplab v3 网络存在正则化调参的过程,该进程也是网络中非常耗时的一个步骤。除此之外,过多的通道数也有可能导致训练速度慢,网络效率偏低等问题。

2.2 PSPNet 网络原理

PSPNet 是针对相对复杂的场景解析问题而提出的网络,解决的是计算机视觉的基本问题。传统的卷积神经网络没有很好地运用场景中上下文的信息,并且一些不明显的类别不易被察觉,于是网络引入金字塔结构来解决这一问题。PSPNet 的网络结构示意图如图 3 所示。

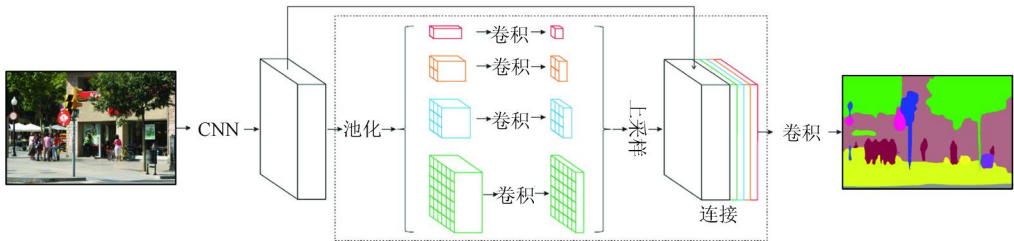


图 3 PSPNet 语义分割网络结构^[20]
Fig. 3 PSPNet semantic segmentation network structure^[20]

PSPNet 的网络结构与 Deeplab v3 网络相似,都是以 ResNet 残差网络为主体骨架,在 ResNet 网络结束后加入各自的设计。网络首先通过预训练的残差网络提取出 feature map 特征图,特征图的大小是输入的 1/8,随后特征图经过金字塔池化模块得到带有整体上下文信息的小图。小图经过上采样后恢复到特征图的尺寸,并且与池化以前的特征图相结合,经过最后一个卷积层后得到最终输出结果。

该金字塔池化模块融合了 4 种不同金字塔尺度的特征信息,这样做有利于帮助网络结合全局的上下文信息。以最上层为例,最上层为最粗糙的 1 * 1 全局池化,生成单个像素多通道输出,后面几层为不同尺度的池化(文中为 2,3,6)。如果金字塔中设定了 N 个级别,在池化以后都要使用 1 * 1 卷积将通道数降为原先的 1/N。卷积后各层次通过双线性插值实现上采样,并和原先的特征图融合在一起。

PSPNet 网络同样存在缺陷。与 Deeplab v3 网络相同,过多的通道数和大量的卷积、池化运算耗

费了大量的时间,影响到了网络的运行效率。PSPNet 网络和 Deeplab v3 网络类似,都将卷积神经网络进行了扩张的操作。相比于传统的卷积神经网络,扩张后的网络虽然参数没有增加太多,但是空间尺寸大得多,因此导致了运行速度的降低。但是相比于 Deeplab v3 网络,PSPNet 网络缺少了正则化的过程。

2.3 ShelfNet 网络原理

ShelfNet 网络是一个同样以 ResNet 网络为骨架的编码 - 解码网络。它和普通的编码 - 解码网络结构的不同点在于,该网络中有很多种编码 - 解码的路径,并且在每一个空间级别上都存在跳跃连接。这种多路径的算法大大提高了网络的精度。ShelfNet 网络的原理图如图 4 所示。图中 A - D 代表了 ResNet 网络中不同的层次；列 1 - 4 代表了不同的分支。为了尽可能地减少通道数量从而提升训练速度,网络在第一列的运算中使用了 1 * 1 卷积层后接以 batch normalization 正则归一以及 ReLU 激活函数,并且将通道的数量减少为 ResNet 网络中的 1/4。

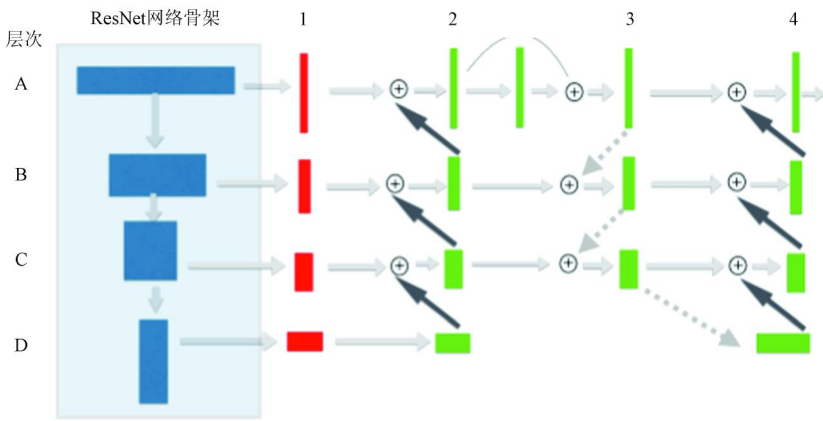


图 4 ShelfNet 语义分割网络结构^[21]

Fig. 4 ShelfNet semantic segmentation network structure^[21]

除此之外,分支 1 和 3 为编码分支;分支 2 和 4 为解码分支。编码器分支中采用跨距为 2 的卷积;在解码器中采用跨距为 2 的转置卷积。从图中可见,该网络最大的特点就是丰富了路径的选择。从输入到输出可以有很多种选择的途径,而不再拘泥于传统编码-解码网络中的单一途径。图 4 中 2~4 列的残差块均为共享权值。共享权值的优点在于结合了跳跃连接优点的同时,比标准残差单元减少了很多参数。它能更有效地提取特征,并且可以加快模型的运行速度。ShelfNet 网络实现了在不同的层次上进行跳转的功能,捕捉了更多的浅层和深层特征,也使得运行速度和分割精度有了显著提升。

3 实验及结果分析

3.1 实验设置及实验流程设计

本文在同一个 Nvidia GeForce GTX 1080ti GPU 显卡上,使用 Deeplab v3 - 50, Deeplab v3 - 101, PSPNet50, PSPNet101, ShelfNet50 和 ShelfNet101 共 6 个网络进行了对比实验。实验的深度学习模型基于 PyTorch 搭建,使用 ResNet 网络作为基础框架。采用公开的 Coco - 2014 数据集和 ImageNet 数据集对网络进行预训练,实验开始时统一将输入图片的大小设置为 256 像素 × 256 像素。循环设置为 80,因为数据集训练到 50 个循环附近就会收敛。初始学习率设为 0.001,学习率会随着训练的深入而下降。

实验第一步是图像裁剪。网络需要有固定尺寸

的输入,故需要将原始图片切割成相应尺寸,在本实验中为长宽为 256 像素 × 256 像素的图像瓦片。除原始的影像外,还需生成与之一一对应的真实值 (ground truth) 标签。

第二步是网络训练。遥感影像每一个像素点的像素值会作为输入传入网络。经过卷积神经网络的运算得到输出,并且和相应的真实值标签进行比较,由损失函数算出损失值。

第三步是返回调参。根据第二步得到的损失函数值的反馈,整个网络框架会返回到卷积的部分调参,对网络整体进行完善。

第四步是结果预测。训练结束以后会得到已完成调参的实验模型。根据模型和测试集进行网络预测,得到最后的预测结果。

3.2 实验数据来源

本实验采用中国深圳市的正射遥感图像作为实验的数据源。选择深圳作为数据集的原因在于该地区城市面积大,易于划分训练集区域,并且在 Google 影像地图中存有不同时相的数据,便于进行深度学习的研究。数据为深圳市 2018 年和 2019 年 Google 影像 18 级高分辨率影像,手工划分出训练集建成区矢量文件作为与影像相对应的标签图。首先通过裁剪得到了 9 577 张 256 像素 × 256 像素大小的影像,其中的 80% 作为训练集,其余作为验证集。验证集的瓦片影像不会参与训练,但是可以起到验证网络精度的作用。数据集集中的部分影像样本如图 5 所示。

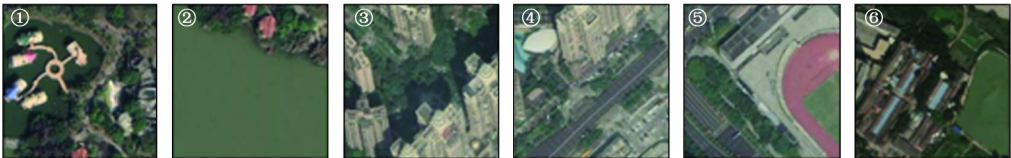


图 5 - 1 部分训练数据集
Fig. 5 - 1 Partial training data set

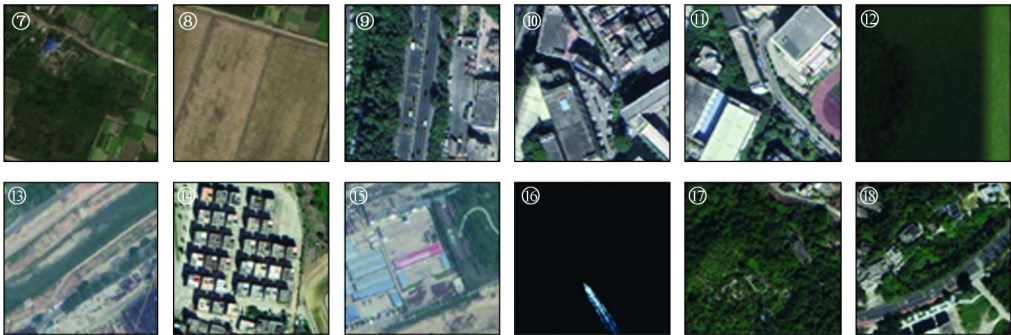


图 5-2 部分训练数据集

Fig. 5-2 Partial training data set

3.3 实验结果与分析

实验中,每个网络均进行 80 个循环,其中每次循环显示一次结果;每 5 次保存一个模型,故最后可以得到 16 个训练不同阶段的模型。实验每一次结果显示均输出 6 个不同参数,分别为:训练集平均损失、验证集平均损失、验证集平均准确率、验证集平均重叠度 (intersection over union, IOU) ($mIOU$)、背景 IOU 和前景 IOU 。为了防止偶然偏大或偏小的结果带来的影响,本实验再选取 2 个评价指标:使用前 20 循环平均前景 IOU 表示网络前期收敛速度;使用 51 – 80 循环平均前景 IOU 来表示网络稳定后的平均 IOU 水准。本实验共 6 个网络参与训练,分别为 Deeplab v3 – 50, Deeplab v3 – 101, PSPNet50, PSPNet101, ShelfNet50, ShelfNet101, 其中数字表示 ResNet 的层数。为了比较网络在前期的收敛速度,实验列出第 20 个循环的 6 个参数,以及

前 20 个循环平均前景 IOU ,结果如表 1 所示;为了比较网络在收敛后的准确度,实验列出最终的 6 个参数,以及 51 – 80 循环平均前景 IOU 和训练时长,结果如表 2 所示。随机森林分类器和支持向量机分类器是遥感影像分类中较经典的机器学习分类器,但由于二者结构与语义分割网络有明显的不同,不能像语义分割模型一样通过模型分类器直接预测得到等尺寸的预测结果图片。所以,在训练这 2 个分类器前需要对影像数据和标签数据分别进行预处理编码成特征向量形式以适应分类器结构。本实验同样针对随机森林分类器和支持向量机分类器进行数据集的预测,但由于二者结构的局限性及特殊性,无法对随机森林分类器和支持向量机分类器进行前期精度的比较以及训练时长的统计。本文只进行最终验证集平均准确率及验证集 $mIOU$ 的横向比较,比较结果如表 2 所示。

表 1 网络前期训练精度比较

Tab. 1 Network early stage training accuracy comparison (%)

网络	训练集平均损失	验证集平均损失	验证集平均准确率	$mIOU$	背景 IOU	前景 IOU	前 20 循环平均前景 IOU
Deeplab v3 – 50	19. 91	12. 15	87. 90	82. 21	94. 30	70. 13	60. 35
Deeplab v3 – 101	18. 92	12. 37	87. 00	81. 65	94. 19	69. 11	58. 05
PSPNet50	19. 21	12. 02	88. 03	82. 35	94. 34	70. 36	63. 30
PSPNet101	18. 33	12. 22	87. 92	82. 27	94. 32	70. 23	63. 27
ShelfNet50	18. 01	12. 48	89. 19	82. 75	94. 33	71. 16	63. 60
ShelfNet101	21. 17	13. 15	86. 82	81. 34	94. 08	68. 60	58. 23

表 2 网络最终训练精度比较

Tab. 2 Network final training accuracy comparison

分类器/网络	训练集平均损失/%	验证集平均损失/%	验证集平均准确率/%	$mIOU$ /%	背景 IOU /%	前景 IOU /%	51 – 80 循环平均前景 IOU /%	训练时长/h
随机森林	—	—	85. 48	55. 98	—	—	—	—
支持向量机	—	—	79. 82	46. 27	—	—	—	—
Deeplab v3 – 50	13. 85	10. 13	90. 65	85. 38	95. 30	75. 46	74. 74	48. 92
Deeplab v3 – 101	12. 78	9. 68	90. 88	85. 70	95. 41	75. 98	75. 68	68. 35
PSPNet50	13. 72	9. 81	90. 78	85. 37	95. 28	75. 45	74. 73	32. 13
PSPNet101	10. 95	8. 66	92. 44	87. 25	95. 88	78. 63	78. 01	51. 86
ShelfNet50	11. 54	9. 08	92. 38	86. 71	95. 65	77. 76	77. 05	14. 02
ShelfNet101	12. 05	9. 31	91. 60	86. 12	95. 49	76. 94	76. 89	15. 63

从表 1 中可以比较得出网络初期的收敛效率。PSPNet50 网络在验证集损失值和背景 *IOU* 上表现更好；而 ShelfNet50 网络在其余的 5 项指标中均为最优。*IOU* 算法为输出结果与真值的交集除以并集,故前景 *IOU* 最可以反映网络的精度。从表 1 的数值中可以得出,ShelfNet50 网络的前期收敛速率是非常理想的,而 Deeplab v3 网络均没有达到很好的效果,这与网络本身的结构有较大的关系；其次,所有以 ResNet50 为骨架的网络在训练前期均比以 ResNet101 为骨架的网络有更好的表现,这是因为 ResNet101 中的参数数量要远多于 ResNet50,故参数少的网络前期收敛更快。

相比于表 1,表 2 则可以体现网络的性能优劣。随机森林分类器和支持向量机分类器虽然在平均准确率方面与语义分割网络相近,但是在 *mIOU* 上却相差甚远。在实验训练时长方面,ShelfNet 网络占据了绝对的优势,并且 ShelfNet 网络的运行速率不会随着 ResNet 网络层级的加深出现过大的变动。相比之下,Deeplab v3 网络和 PSPNet 的 4 个网络在训练上花费了过多时间。这说明了在讲究网络的运行速率时,ShelfNet 网络是绝对的最优选。

在准确率指标中, PSPNet101 网络拥有最高的精度,其次是 ShelfNet50 网络和 ShelfNet101 网络,而 Deeplab v3 网络在最终的准确率一项同样表现不佳。在验证集前景 *IOU* 一项中, PSPNet101 网络能够达到 78.01% 的精度,比 ShelfNet50 网络高出 1 个

百分点,这说明 PSPNet101 网络在分割建成区的领域可达到非常优异的效果。值得注意的是, ShelfNet101 网络的各项指标低于 ShelfNet50 网络,说明网络可能在训练时出现了过拟合。

综上, PSPNet101 网络的分割精度最优,但是考虑到 PSPNet101 网络的训练时长过长,综合各种指标可以得出: ShelfNet50 网络是在识别建成区中综合效率最高的选择。

由实验结果回推至理论本身, ShelfNet 网络正是拥有了共享权值以及可以在不同空间层次上进行跳转连接的功能,使其比标准残差单元减少了很多参数。实验证明,这些改变可以显著加快模型的运行速度。除此之外, ShelfNet 网络最大的特点就是丰富了路径的选择。正是这一个特点使得网络能够学习到各种浅层和深层的信息,从而保证了网络优秀的性能。

相比之下, PSPNet 和 Deeplab v3 网络在卷积运算时空间尺寸大得多,因此导致了运行速度的降低。而相比于 PSPNet 网络, Deeplab v3 网络在编码过程结束后直接解码到原尺寸,从而减少了信息的读取。实验结果表明, Deeplab v3 网络的指标确实稍逊于其他 2 种网络。这可能是因为网络本身的缺陷,同样可能是因为 Deeplab v3 网络不适用于上下文联系较强的识别任务,而专长于另一些种类的识别任务。

本文使用 ShelfNet50 网络识别 2019 年深圳市 18 级遥感影像,结果如图 6 所示。



图 6-1 ShelfNet50 网络测试集分割结果

Fig. 6-1 ShelfNet50 network test set segmentation results

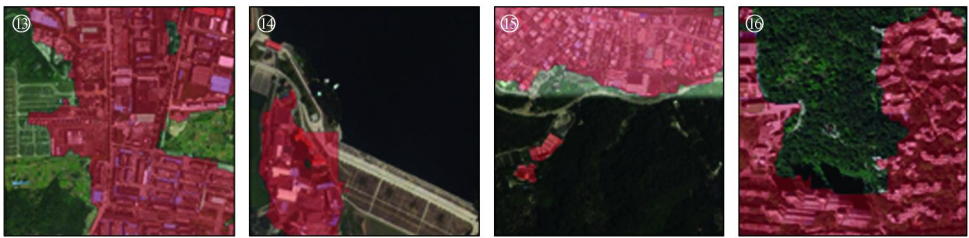


图 6-2 ShelfNet50 网络测试集分割结果

Fig. 6-2 ShelfNet50 network test set segmentation results

从图 6 中可见,其中红色为建成区,除去边界上的细微识别误差,ShelfNet50 网络可以很好地识别深圳市 2019 年遥感影像中各种不同类别的建成区。预测结果同样表明,ShelfNet50 网络有较好的上下文推理能力。城市小区的绿化和乡村大片的植被地物特征完全相同,而 ShelfNet50 网络能够从周围的

地物特征推断该地区的具体类别,这也体现出了空洞卷积的重要性。实验表明,ShelfNet50 网络可以适用于遥感影像城市建成区的识别,并且拥有很好的识别效果。

然而,ShelfNet50 网络在识别时同样存在误差,如图 7 所示。

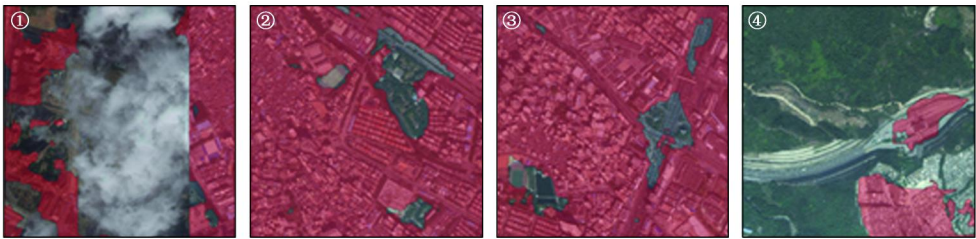


图 7 ShelfNet50 网络测试集分割误差

Fig. 7 ShelfNet50 network test set segmentation error

图 7 中反映出 3 个 ShelfNet50 网络的问题。首先,当原始影像有云时,网络不能判断云下的地物,这也是所有语义分割网络共同的缺陷;其次,对于城区内一些大型非建筑物(公园绿地、大型立交桥),ShelfNet50 网络还是欠缺识别此类地物的能力,而这也与空洞卷积核的大小有关;最后,小部分建筑物存在漏识别的现象,这也说明了训练集中缺少该类标签,此类问题可以通过扩充数据集的方法解决。

4 结论

本文基于多种深度学习语义分割方法,针对高分遥感影像城市建成区提取问题做了深入的比较研究。本文首先从网络的基本原理出发,深入分析并比较了 Deeplab v3 网络、PSPNet 网络以及 ShelfNet 网络这 3 种均以 ResNet 残差网络为骨架的语义分割网络的结构差别,并通过网络结构分析了每一个网络存在的优缺点。其次,本文通过控制变量法,使用同一套实验装置以及同一个数据集对 6 种语义分割网络进行了测试及结果分析。本文通过实验得出的结果返回原理分析部分,验证之前分析的合理性。最后,实验针对综合表现最佳的 ShelfNet50 网络进

行结果预测,印证语义分割网络在高分遥感影像城市建成区分割中的可行性。综合以上,本文得到以下结论:

- 1)实验的初始收敛阶段,ResNet50 残差网络的表现普遍好于 ResNet101 网络,其中 ShelfNet50 网络的收敛效果最为优异。
- 2)在网络最终分割效果的评定中,PSPNet 的各项指标均处于 6 个语义分割网络的第 1 名;验证集前景 *IOU* 比 ShelfNet50 网络高出 1 个百分点。
- 3)运行时长方面,ShelfNet 网络远远短于另外 2 类网络,且 ShelfNet 网络基本不受网络层数的影响。综合网络精度以及运行速度 2 方面因素综合考虑,ShelfNet50 网络是解决高分遥感影像城市建成区识别的最优网络。
- 4)ShelfNet50 网络在最终的预测过程中表现优异,具有良好的识别效果,说明 ShelfNet50 网络可以完美解决城市建成区的识别问题。

参考文献 (References):

[1] 中华人民共和国建设部. 城市规划基本术语标准: GB/T50280—1998[S]. 北京:工程建设标准全文信息系统,1999. Ministry of Construction of the People's Republic of China. Basic terminology standard for urban planning: GB/T50280—1998 [S]. Beijing: Engineering Construction Standard Full - Text In-

- formation System, 1999.
- [2] Sinha P, Verma N K, Ayele E. Urban built - up area extraction and change detection of Adama municipal area using time - series Landsat images [J]. International Journal of Advanced Remote Sensing and GIS, 2016, 5(8): 1886 - 1895.
- [3] Zha Y, Gao J, Ni S. Use of normalized difference built - up index in automatically mapping urban areas from TM imagery [J]. International Journal of Remote Sensing, 2003, 24(3): 583 - 594.
- [4] 陈志强, 陈健飞. 基于 NDBI 指数法的城镇用地影像识别分析与制图 [J]. 地球信息科学学报, 2012, 8(2): 137 - 140.
Chen Z Q, Chen J F. Urban land image recognition analysis and mapping based on NDBI index method [J]. Geo - Information Science, 2012, 8(2): 137 - 140.
- [5] 梁珊, 刘荣, 祖帅, 等. 基于多光谱遥感影像的城市建设用地提取及扩张分析 [J]. 测绘与空间地理信息, 2018, 41(6): 74 - 78.
Liang S, Liu R, Zu S, et al. Urban construction land extraction and expansion analysis based on multi - spectral remote sensing images [J]. Geomatics and Spatial Geographic Information, 2018, 41(6): 74 - 78.
- [6] Liaw A, Wiener M. Classification and regression by random forest [J]. R News, 2002, 2(3): 18 - 22.
- [7] Burges C J C. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121 - 167.
- [8] Gowthami K, Thilagavathi K. Built - up area detection of remote sensing images using static clustering technique [C]//2014 International Conference on Electronics and Communication Systems (ICECS). IEEE, 2014: 1 - 5.
- [9] Gong P, Liu H, Zhang M, et al. STab: classification with limited sample; Transferring a 30 - m resolution sample set collected in 2015 to mapping 10 - m resolution global land cover in 2017 [J]. Science Bulletin, 2019, 64(6): 370 - 373.
- [10] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541 - 551.
- [11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems, 2012: 1097 - 1105.
- [12] 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): 259 - 266.
Yang J Y, Zhou Z X, Du Z R, et al. High - resolution remote sensing imagery rural construction land extraction based on SegNet semantic model [J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(5): 259 - 266.
- [13] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder - decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481 - 2495.
- [14] 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 [J]. 计算机工程与应用, 2019, 55(7): 212 - 218.
Su J M, Yang L X, Jing W P. High - resolution remote sensing image semantic segmentation method based on U - Net [J]. Computer Engineering and Applications, 2019, 55(7): 212 - 218.
- [15] 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. Springer, Cham. Berlin, German: Springer, 2015: 234 - 241.
- [16] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. 12017 - 12 - 05. <http://arxiv.org/odf/1706.05587.pdf>.
- [17] Yu F, Koltun V. Multi - scale context aggregation by dilated convolutions [J/OL]. [2019 - 05 - 30]. <https://arxiv.org/abs/1511.07122.pdf>.
- [18] 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.
Wang J Q, Li J S, Zhou H C, et al. Typical elements extraction method of remote sensing image based on Deeplabv3 + and CRF [J]. Computer Engineering, 2019, 45(10): 260 - 265, 271.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770 - 778.
- [20] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2881 - 2890.
- [21] Zhuang J T, Yang J L, Gu L, et al. ShelfNet for real - time semantic segmentation [EB/OL]. (2018 - 11 - 27). <http://arxiv.org/pdf/1811.11254.v1.pdf>.
- [22] Simonyan K, Zisserman A. Very deep convolutional networks for large - scale image recognition [EB/OL]. (2014 - 09 - 04). <http://arxiv.org/pdf/1409.1556.v1.pdf>.

Research and comparative analysis on urban built - up area extraction methods from high - resolution remote sensing image based on semantic segmentation network

LIU Zhao, ZHAO Tong, LIAO Feifan, LI Shuai, LI Haiyang
(Institute of Transportation Engineering and Geomatics, Department of Civil Engineering, Tsinghua University, Beijing 100084, China)

Abstract: The extraction of urban built - up areas plays an important role in urban development planning. To find

out the method of extracting remote sensing image urban built - up area based on convolutional neural network which can balance efficiency and recognition accuracy , the authors started with the principle of neural network structure and compared as well as analyzed the internal structure of multiple semantic segmentation networks . The semantic segmentation network was trained separately and the results were comparatively studied . The experimental result shows that the ShelfNet - 50 network could ensure high recognition accuracy while training speed , achieved 77% foreground segmentation accuracy while training time was only 14 hours , and the result of ShelfNet - 50 network prediction was also highly consistent with the corresponding remote sensing image data . The experiment confirms that ShelfNet - 50 network can be applied to high - resolution remote sensing image urban built - up area extraction problems.

Keywords: high - resolution remote sensing image ; convolutional neural network ; semantic segmentation ; urban built - up area

(责任编辑: 陈 理)

下期要目

- 汪 雅 一种近景倾斜摄影的三维建模盲区自动修复技术
- 陈 栋 基于 Arcpy 与定制 ArcToolbox 的矿山新增图斑自动编号及方法改进
- 宋 奇 阿拉尔垦区近三十年耕地变化及其驱动因子分析
- 安健健 基于 Faster R - CNN 的火电厂冷却塔检测及工作状态判定
- 陈宝林 历史平均值法用于 MODIS 影像像元云补偿——以甘肃省为例
- 郝固状 红崖山水库近 20 年面积变化遥感调查及驱动力分析
- 叶婉桐 基于 GEE 的近 20 年升金湖湿地不同季节地表温度时空变化及地表类型响应
- 牟晓莉 Google Earth Engine 在土地覆被遥感信息提取中的研究进展
- 李根军 ZY1 - 02D 高光谱载荷数据在地质矿产调查中的应用分析
- 徐甜雨 基于多元分区建模指标优化的“一带一路”人口空间化研究
- 赵龙贤 基于 RS 和 GIS 技术的西藏多龙矿集区矿山选址研究
- 凌 晓 基于致灾因子对称法分级的信息量模型在地震滑坡危险性评价中的应用
- 张 腾 利用 Sentinel - 1 和 ALOS - 2 数据探测茂县滑坡
- 韦 耿 基于精细模式气溶胶与 WRF 模式估算 PM2.5 质量浓度
- 崔鹏霞 基于残差网络特征融合的高光谱图像分类
- 吴 霞 基于多光谱遥感的盐渍化评价指数对宁夏银北灌区土壤盐度预测的适用性分析
- 赵 怡 基于 Sentinel 2A 和 Landsat 8 的城市不透水面的提取
- 周 燕 基于 GOCI 数据的胶州湾水体透明度遥感反演及日变化研究
- 李志勇 一种基于控制点的影像内部几何精度检测方法