

doi: 10.6046/gtzyyg.2020.04.12

引用格式: 刘钊,廖斐凡,赵桐. 基于 PSPNet 的遥感影像城市建成区提取及其优化方法[J]. 国土资源遥感,2020,32(4):84-89. (Liu Z, Liao F F, Zhao T. Remote sensing image urban built-up area extraction and optimization method based on PSPNet [J]. Remote Sensing for Land and Resources, 2020, 32(4): 84-89.)

基于 PSPNet 的遥感影像城市建成区提取及其优化方法

刘 钊, 廖斐凡, 赵 桐

(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: 利用高分辨率卫星遥感影像提取建成区边界对于城市扩张监测和城市发展规划具有重要意义。为获取高精度高空间分辨率的建成区数据,本研究通过归一化建筑指数(normalized difference built-up index, NDBI)加入人工目视解译方法构建城市建成区遥感影像数据集,分别采用传统机器学习方法和包括 PSPNet 在内的 4 种深度学习语义分割网络对 Sentinel-2 影像进行建成区提取,训练结果表明 PSPNet 网络对于建成区的提取具有最高的精度(训练集交并集比(intersection over union, IOU)为 79.5%)。提出 Overlapsize 方法对 PSPNet 的提取结果进行优化,进一步提高了建成区提取准确率,该方法在训练集上的 IOU 达到 80.5%,在测试集上的 IOU 达到了 83.1%,利用 PSPNet + Overlapsize 提取建成区的方法相较于传统机器学习方法具有实际应用意义。

关键词: 建成区提取; 深度学习; 卷积神经网络; 语义分割; PSPNet; Overlapsize

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 1001-070X(2020)04-0084-06

0 引言

建成区在《城市规划基本术语标准》^[1]中的定义为:“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。”一个城市的建成区边界相较于行政界更能反映一个城市的发展规模和程度,高精度、高质量、高实时性的城市建成区数据是研究城市空间格局变化、城市扩张变迁、土地资源管理和产业布局的关键。因此研究建成区提取方法是一个具有重要现实意义的课题。

随着遥感技术的发展,国内外有很多学者围绕遥感影像建成区提取展开了研究,所提出的方法多种多样,2000 年 Masek 等^[2]就已经通过对华盛顿地区不同年份的 MSS 和 TM 影像采用归一化植被指数(normalized difference vegetation index, NDVI)插值法,得到了 1973—1996 年间城市扩展区域的变化,但这一方法只适用于植被覆盖较好的地区,不具有普遍适用性;刘智丽等^[3]根据目标地物波谱特征,应用归一化建筑指数(normalized difference built-up index, NDBI)结合 NPP-VIIRS 指数进行城市建成区提取也得到了精度较高的结果,但这一类指数方法也不可避免地存在同谱异物和同物异谱误识别的现象,会影响识别精度。而近年来机器学习方法如随机森林(random

forest, RF)^[4]和支持向量机(support vector machine, SVM)^[5]等也被运用在建成区的提取实验中,这一类方法更加注重地物的纹理特征和空间特征,避免了传统指数方法的缺陷,提取效果更好,但是这一类方法也存在训练样本数量局限,模型结构简单,分类精度不够且提取结果的空间分辨率不高的问题。

近年来很多学者尝试运用深度学习中的卷积神经网络来提取建成区边界,冯丽英^[6]在基于深度学习技术的高分辨率遥感影像建设用地信息提取研究中,为了解决影像信息提取的边界性问题,以不同尺度对高分二号影像进行分割,然后采用神经网络进行提取,根据结果确定最佳的分割尺度,取得了较 SVM 和 RF 等方法更为优异的结果;马凯等^[7]将卷积神经网络 GoogleNet Inception 网络应用于 Landsat8 影像中进行青海湖地区的用地分类,得到人工用地信息;陈磊士等^[8]直接将 GoogleNet Inception V3 网络用于城市建设用地的提取,取得了比 SVM 方法更优的效果。但以上方法只是针对部分特定地区,模型不具备泛化能力,也没有完全解决边界拼合问题,且提取结果空间分辨率也有待进一步提高。

为进一步提高建成区的提取空间分辨率和精度,获取一个具有普适性的模型,本研究拟建立我国胡焕庸线以东范围内城市建成区 Sentinel-2 遥感影像数据集,采用改进的深度学习卷积神经语义分

割网络(pyramid scene parsing network ,PSPNet)进行建成区提取,并设置多组对比实验以验证方法有效性。在此基础上,为解决边界问题,应用 Overlapsize 方法在模型预测阶段进行改进,探讨这一方法对提取结果准确率的影响。

1 原理与方法

1.1 PSPNet 及其结构特点

PSPNet^[9]是在全卷积网络^[10](fully convolutional network ,FCN)基础上改进而来的语义分割网络,其核心思想是当在对于局部小目标进行判别时,如果在分割层引入更多的全局信息,可以相对降低误识别的概率。本研究中 PSPNet 实现全局信息引入的方

式有 2 种:一是在 PSPNet 的 Resnet^[11]卷积模块中运用空洞卷积增加感受野;二是通过应用金字塔池化模块分别提取图像深层和浅层的特征并将其融合以降低误分割的概率。金字塔池化模块的运用是 PSPNet 区别于其他网络的主要结构特征。

PSPNet 主要结构大致可以分为特征提取的 Resnet 卷积模块、金字塔池化模块和最后的 FCN 输出模块 3 个部分(图 1)。①输入图像首先通过深度为 50 层的 Resnet 卷积模块提取得到图像特征层;②将所得的图像特征层输入金字塔池化模块来获取影像深层和浅层(池化特征的尺寸分别为 1,2,3,6)特征;③在金字塔池化模块后,由步骤②获得的深层和浅层特征图层会一起输入 FCN 模块最后得到与输入图像相同尺寸的预测图像。

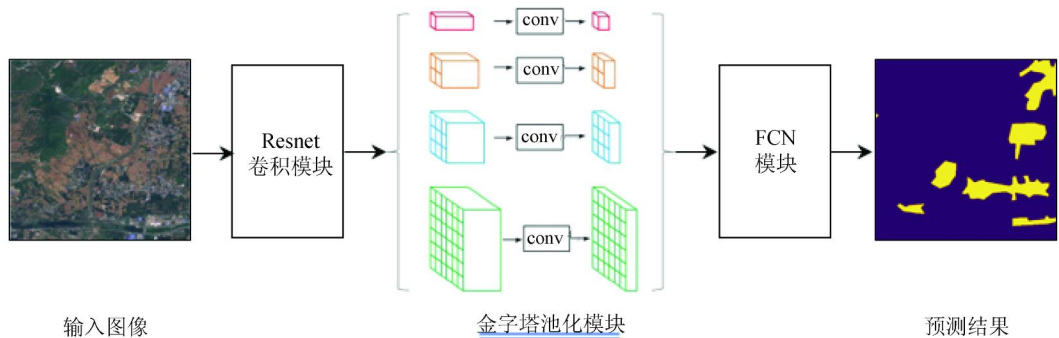


图 1 PSPNet 基本结构

Fig. 1 Basic structure of PSPNet

1.2 应用 PSPNet 模型的遥感影像建成区提取方法

应用 PSPNet 模型进行遥感影像建成区提取的

方法主要分为 2 个部分,模型训练和应用模型预测提取。训练提取流程如图 2 所示。

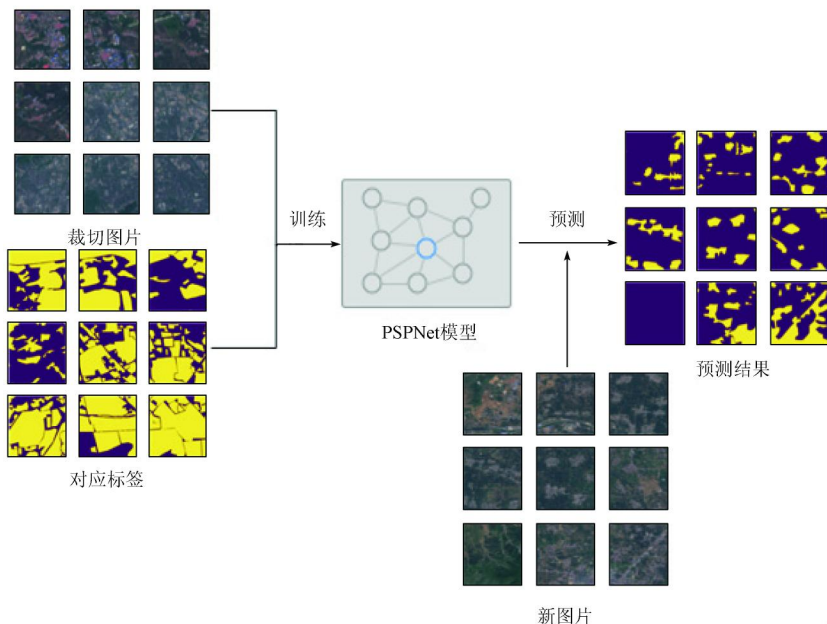


图 2 训练提取流程

Fig. 2 Training extraction flow chart

在模型训练阶段,首先对 Sentinel - 2 影像进行

裁切,建立建成区遥感影像数据集,将所得的数据集

输入至 PSPNet 模型中进行特征学习,得到初步的预测概率图,然后采用交叉熵^[12]计算预测结果与样本真值之间的损失值,并利用 Adam 算法^[13]进行反向迭代优化,在设置好模型训练参数后进行多轮训练直至预测损失小至一定阈值且预测精度变化收敛为止,得到最优模型;在影像预测阶段,将新的 Sentinel-2 影像输入已训练好的 PSPNet 模型进行预测,最终得到高精度的提取结果。

2 实验过程

2.1 遥感影像数据区域选取

本研究选取了我国胡焕庸线以东各地有典型地域特点的城市(哈尔滨、北京、保定、德州、杭州、无锡、武汉、广州、珠海)局部市区及郊区影像作为训练数据集的目标区,目的是获得一个能够适用于全国大部分城市的提取模型。另外,选取异于以上地区的其他城市部分区域(大连、天津、太原、上海、宁波、长沙、昆明)作为独立测试集,用以验证方法的适用性。

2.2 影像数据准备

本研究所采用的影像数据为 Sentinel-2,该影像数据是现在能从公开免费渠道获取的精度最高数据,其中包含 13 个波段的多光谱影像,光谱范围涵盖可见光、近红外与短波红外等。Sentinel-2 影像宽度为 290 km,空间分辨率最高为 10 m,拍摄范围可以覆盖我国全境。

本文采用 2016—2018 年间各目标区 4—10 月份的无云影像,选取常规的 RGB 3 个波段影像作为训练影像数据,一方面是其包涵的纹理信息最为丰富,符合人眼视觉习惯,方便通过目视解译标注训练标签;另一方面是由于下载的影像产品中已经存在的 TCI 影像就是经过大气校正处理后生成的真彩色合成影像,影像质量较好,可以直接使用。

2.3 标签的定义

建成区划定的方法通常有 2 类^[14],一类是城乡规划部门通过已有累计规划管理数据进行划定,这一判定方法主要从区分是否开发建设、是否利用城市基础设施和服务设施的角度出发,主要是关注用地的开发和功能属性;另一类是通过卫星或遥感影像进行判定,将遥感影像上分布有建筑物的区域近似的认为是城市建成区,这种方法主要根据所判别对象在影像上表现出来的颜色和纹理特征来进行分类。二者有一定的区别,本文采用的是第二类划定方法。

依据李爱民^[15]对于建成区边界界定,绝大多数建成区在遥感影像上有以下几个特点:①建成区是一类面积较大的组合型目标群体;②其内部有建筑

物、广场、街道、植被、水域组成;③建成区建筑物数量多,一般以群的形式出现,且彼此间距离很近,排列紧致有序,其内部被街道纵横,被分为若干个街区;④建成区外部多为田野、山地以及小面积居民地,城市之间通过铁路或公路连接;⑤城市建成区内的主要成分建筑物街道、广场一般为水泥表面,即“不透水面”,在遥感影像中光谱特性类似;⑥城市区域内地物目标种类复杂,内部影像灰度变化较为剧烈,而区域周围由于地物单一往往灰度变化较为平缓,建成区与非建成区在绝大多数情况下有着比较明显的边界。

本研究在利用遥感影像标定城市建成区范围时,充分考虑了以上特性,并为了兼顾建成区的空间连续性,将城市内部面积较小的非城市功能用地、水体和绿地也算入城市建成区,另外由于建成区在遥感影像上的纹理特征与城市外散布的大小居民地(县城、集镇、村庄)基本一致,在划分样本时也将居民地划入建成区,以提高模型训练精度。这里特别说明,定义的建成区标签有别于城市规划部门所定义的建成区。据此本研究应用 NDBI^[16]指数界定城市建成区边界的方法初步提取建成区边界,再根据建成区的影像特征通过人工目视检查对数据进行编辑修正,最终获取与影像严格对应的标签数据。

2.4 数据的裁切处理

在将与影像对应的标签数据栅格化后,考虑到将整景遥感影像输入到模型中训练时所占用资源大大超过了机器的负荷,本研究将参与训练的遥感影像和标签裁剪成每张 512 像素 × 512 像素的小图像,最后得到 3 058 对图像及其对应标签,并将其以 8:2 的比例随机分为训练集与验证集。训练集中部分城市影像及其标注数据如图 3 所示(影像上的黄色掩模部分为建成区标注数据)。

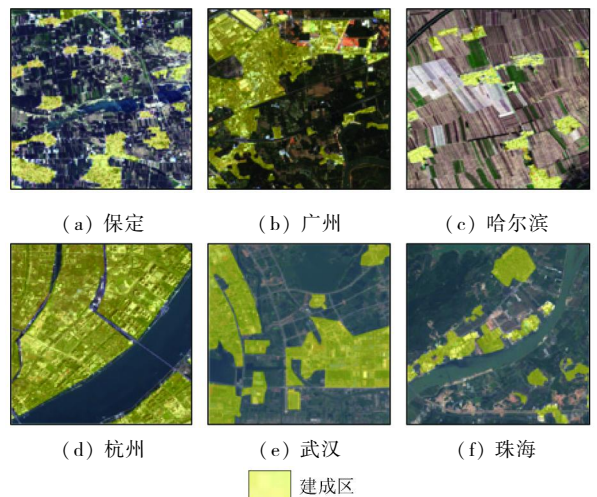


图 3 部分城市影像及其标注数据

Fig. 3 Partial city images and annotated data

2.5 PSPNet 网络训练

本研究中的训练应用 python 的 pytorch 深度学习框架实现,其详细结构见文献[17]。网络分类设置为 2 (建成区和非建成区 2 类),初始学习率设置为 0.001,迭代次数为 50,当迭代训练验证集精度稳定时学习率降为原来的 1/10,最小学习率为 0.000 001。经试验证明此时验证集精度已经收敛不再有显著提高。

为了评估分类精度,采取总体精度(overall accuracy, OA)和交并集比(intersection over union, IOU)作为评价指标。OA 为总体预测正确的像素数占总像素数的比例,IOU 为建成区真实值和预测值 2 个集合交集与并集之比。2 个指标的计算公式分别为:

$$OA = \frac{\sum_{i=0}^1 p_{ii}}{\sum_{i=0}^1 \sum_{j=0}^1 p_{ij}}, \quad (1)$$

$$IOU = \frac{p_{11}}{p_{01} + p_{11} + p_{10}}, \quad (2)$$

式中 p_{ij} 为本属于 i 类被分为 j 类的像素数量,其中 1 代表建成区,0 代表非建成区。

3 结果与优化

3.1 训练结果

为了验证本文方法的有效性,将分别应用 RF, SVM 以及卷积神经网络中的 Deeplab - V3^[18], DlinkNet^[19] 和 ShelfNet^[20] 等模型对比验证 PSPNet 模型在建成区提取方面的适用性,经过训练后,得到不同模型方法在验证集上的 OA 和 IOU 得分,如图 4 所示。由训练结果可以看出,深度学习的方法较机器学习的方法在 IOU 上至少提高了 21 百分点,而 PSPNet 网络在训练中分别得到最高的 OA(96.7%)和 IOU(77.3%)。

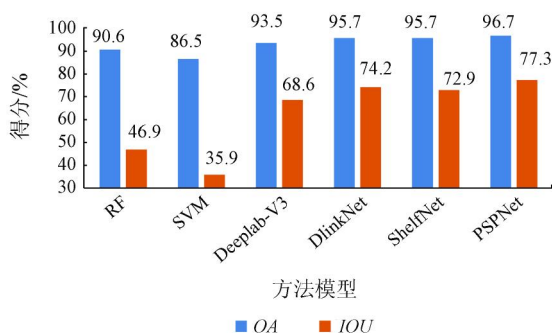


图 4 训练结果

Fig. 4 Training results

3.2 结果优化

应用 PSPNet 网络模型进行预测虽然能够得到比较好的结果,但从实际效果观察来看,还有明显的问题,部分预测结果有明显的不连续性,在裁切线边缘有断层现象,这明显不符合城市建成区空间分布逻辑,属于明显错误。初步分析是因为最后得到的提取结果是由各个分割小图像提取结果合并得到的,而当被预测对象在图像边缘被相邻图像分割成 2 部分时由于信息不完整误识别率较高,进而出现被相邻 2 幅图像分割的对象在一侧图像被识别而在另一侧没有被识别的情况,这就导致在最终的合并结果中出现了断层。

基于以上分析,本研究提出了 Overlapsize 方法来验证并解决这一问题,具体思路是,在应用模型进行提取时,裁剪影像采取冗余分割,即相邻图像在分割的时候会在基准尺寸(512 像素 × 512 像素)的基础上再外扩一定范围(128 像素),得到与相邻图像存在重叠区域的外扩图像(768 像素 × 768 像素),在将每一外扩图像输入模型得到提取结果之后,将提取结果裁剪去掉外扩部分保留基准尺寸部分再进行拼接。其流程图如图 5 所示,对比应用 Overlapsize 方法前后建成区提取结果如图 6 所示。

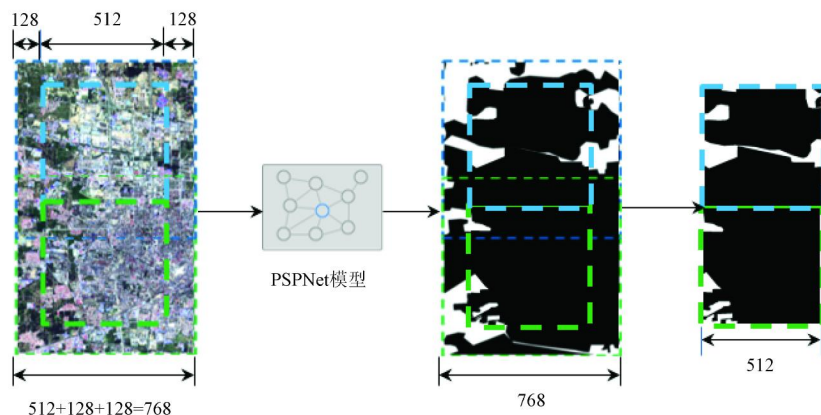


图 5 Overlapsize 方法示意

Fig. 5 Method schematic of Overlapsize

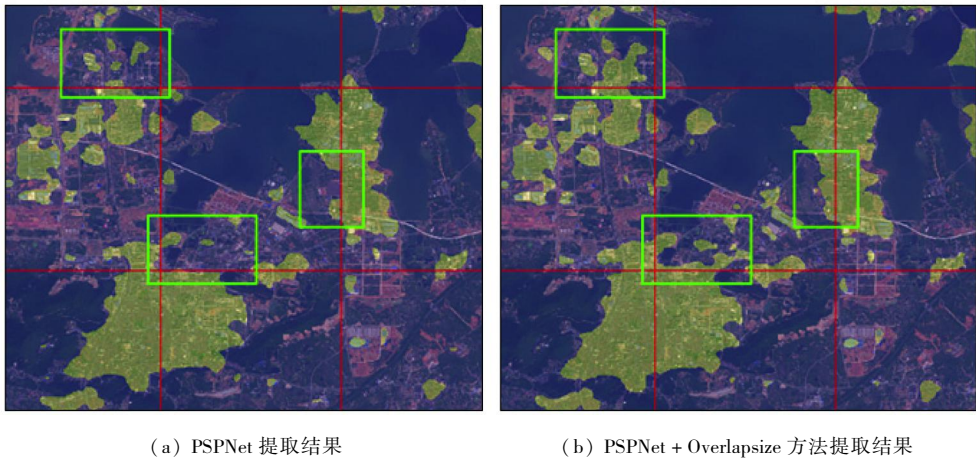


图 6 应用 Overlapsize 方法前后建成区提取结果

Fig. 6 Extraction results of built - up area before and after using Overlapsize

图 6 中红线为裁切线。可以看出,图 6(a)部分预测结果有明显的 discontinuity,在裁切线边缘(绿框区域)有断层现象,图 6(b)的对应区域则平滑且自然,说明 Overlapsize 方法可以有效地解决相邻预测结果在拼接时产生的边界断层问题,而检测结果也证明在 PSPNet 模型中运用 Overlapsize 方法后训练集和测试集 *IOU* 分别提高了 2.2 和 2.6 百分点,如图 7 所示。

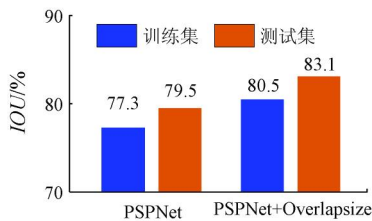


图 7 改进效果对比

Fig. 7 Comparison of improvement effects

4 结 论

本文面向 Sentinel - 2 高分辨率遥感影像,通过 NDBI 和人工目视解译修正的方法建立了大规模的遥感影像数据集,对数据集采用几种经典机器学习方法和神经网络方法进行训练,并将其中最优的学习模型利用 Overlapsize 方法进行优化,通过实验主要得到以下结论:

1) 几种经典语义分割神经网络在提取建成区时较传统机器学习方法如 SVM 和 RF 方法有更高的精度和适用性。其中 PSPNet 的表现最好,训练集 *IOU* 为 77.3%,测试集 *IOU* 为 79.5%,具有良好的适用性。

2) 利用 Overlapsize 优化方法能够有效地解决预测结果边界断层问题,并进一步提高预测精度。利用 PSPNet + Overlapsize 方法提取建成区具有较好的应用预期。

3) 本研究训练模型对于影像中蔬菜大棚和机场存在误识别的现象,这是训练样本缺失造成的,下一步工作中还需要补充样本数据继续迭代训练完善模型。

参考文献 (References):

[1] 中华人民共和国建设部. GB/T50280—98 城市规划基本术语标准[S]. 工程建设标准全文信息系统,1999.
Ministry of Construction P. R. China. GB/T50280—98 standard for basic terminology of urban planning[S]. Engineering Construction Standard Full - Text Information System, 1999.

[2] Masek J G, Lindsay F E, Goward S N. Dynamics of urban growth in the Washington DC metropolitan area, 1973—1996, from Landsat observations[J]. International Journal of Remote Sensing, 2000, 21 (18): 3473 - 3486.

[3] 刘智丽,张启斌,岳德鹏,等. 基于 Sentinel - 2A 与 NPP - VIIRS 夜间灯光数据的城市建成区提取[J]. 国土资源遥感, 2019, 31 (4): 227 - 234. doi:10.6046/gtzyyg. 2019. 04. 29.
Liu Z L, Zhang Q B, Yue D P, et al. Extraction of urban built - up areas based on Sentinel - 2A and NPP - VIIRS nighttime light data [J]. Remote Sensing for Land and Resource, 2019, 31 (4): 227 - 234. doi:10.6046/gtzyyg. 2019. 04. 29.

[4] 于清永. 利用中高分辨率影像对城市建成区信息提取的方法研究[D]. 阜新: 辽宁工程技术大学, 2016.
Yu Q Y. Research on information extraction of urban built - up area by using medium and high resolution images [D]. Fuxin: Liaoning Technical University, 2016.

[5] 闫晓天. 基于支持向量机及 MODIS 数据的南昌市城市空间格局演变研究[D]. 南昌: 东华理工大学, 2016.
Yan X T. Study on evolution of urban spatial pattern of Nanchang base on support vector machine and MODIS data [D]. Nanchang: East China University of Technology, 2016.

[6] 冯丽英. 基于深度学习技术的高分辨率遥感影像建设用地信息提取研究[D]. 杭州: 浙江大学, 2017.
Feng L Y. Research on construction land information extraction from high resolution images with deep learning technology [D]. Hangzhou: Zhejiang University, 2017.

[7] 马凯, 罗泽. 基于卷积神经网络的青海湖区域遥感影像分类[J]. 计算机系统应用, 2018, 27(9): 137 - 142.

- Ma K, Luo Z. Classification of remote sensing images in Qinghai Lake based on convolutional neural network [J]. *Computer Systems and Applications*, 2018, 27(9): 137 – 142.
- [8] 陈磊士, 赵俊三, 董智文, 等. 基于深度学习的滇中城市多光谱影像建设用地信息提取 [J]. *软件导刊*, 2018, 17(11): 177 – 180, 186.
- Chen L S, Zhao J S, Dong Z W, et al. Urban construction land information extraction based on deep learning by multi-spectral remote sensing imagery of Yunnan central urban agglomeration area [J]. *Software Guide*, 2018, 17(11): 177 – 180, 186.
- [9] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 2881 – 2890.
- [10] Jonathan L, Evan S, Trevor D. Fully convolutional networks for semantic segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 3431 – 3440.
- [11] 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*. Las Vegas: IEEE, 2016: 770 – 778.
- [12] Wang T, He P L. The classification algorithm based on the cross entropy rule and new activation function in fuzzy neural network [C]// *2005 International Conference on Machine Learning and Cybernetics*. Guangzhou: IEEE, 2005, 8: 4631 – 4635.
- [13] Kingma D P, Ba J. Adam: A method for stochastic optimization [EB/OL]. (2014 – 12 – 22) [2019 – 10 – 01]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [14] 肖莹光. 从部门协调的角度看建成区和规划建设用地的定义和划定 [C]// *中国城市规划学会、重庆市人民政府. 规划创新: 2010 中国城市规划年会论文集*. 中国城市规划学会、重庆市人民政府: 中国城市规划学会, 2010: 120 – 124.
- Xiao Y G. Definition and delineation of built-up areas and planned construction land from the perspective of department coordination [C]// *Urban Planning Society of China, Chongqing Municipal Government. Planning and Innovation: Essays of 2010 Annual National Planning Conference*. Urban Planning Society of China, Chongqing Municipal Government: Urban Planning Society of China, 2010: 120 – 124.
- [15] 李爱民. 基于遥感影像的城市建成区扩张与用地规模研究 [D]. 郑州: 解放军信息工程大学, 2009.
- Li A M. Research on urban built-up area expansion and land use scale based on remote sensing [D]. Zhengzhou: Information Engineering University, 2009.
- [16] Zha Y, Gao J, Ni S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery [J]. *Remote Sensing*, 2003, 24(3): 583 – 594.
- [17] Zhang H, Dana K, Shi J, et al. Context encoding for semantic segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 7151 – 7160.
- [18] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. (2017 – 06 – 17) [2019 – 06 – 01]. <https://arxiv.org/pdf/1706.05587.pdf>.
- [19] Zhou L C, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 182 – 186.
- [20] Zhuang J, Yang J L, Gu L, et al. ShelfNet for fast semantic segmentation [EB/OL]. (2019 – 09 – 04) [2019 – 10 – 01]. <https://arxiv.org/pdf/1811.11254.pdf>.

Remote sensing image urban built-up area extraction and optimization method based on PSPNet

LIU Zhao, LIAO Feifan, ZHAO Tong

(*Institute of Transportation Engineering and Geospatial Information, Department of Civil Engineering, Tsinghua University, Beijing 100084, China*)

Abstract: Using high-resolution satellite remote sensing images to extract the boundary of the built-up area is of great significance for urban expansion monitoring and urban development planning. In order to obtain high-precision and high-resolution built-up area data, this study uses the NDBI index and artificial visual interpretation methods to construct remote sensing image datasets of urban built-up areas and uses traditional machine learning methods and four deep learning methods including PSPNet semantic segmentation network to extract the built-up area of Sentinel-2 images. The training results show that the PSPNet network has the highest accuracy for the built-up area extraction (IOU of the training set is 79.5%). This paper employs Overlapsize method to optimize the extraction results of PSPNet, which further improves the accuracy of the built-up area extraction. The IOU on the training set reaches 80.5%, and the IOU on the test set reaches 83.1%. Compared with the traditional machine learning method, the method of PSPNet + Overlapsize has practical application significance in built-up area extracting.

Keywords: built-up area extraction; deep learning; convolutional neural network; semantic segmentation; PSPNet; Overlapsize