



引文格式:王严,刘万军,谭亚丽,等.卷积神经网络滑窗识别结合语义推理的城市功能区分类[J].武汉大学学报(信息科学版),2023,48(6):950-959.DOI:10.13203/j.whugis20210377

Citation: WANG Yan, LIU Wanjun, TAN Yali, et al. Classification of Urban Functional Areas by Convolution Neural Network Recognition Combined with Sliding Window and Semantic Reasoning[J]. Geomatics and Information Science of Wuhan University, 2023, 48(6):950-959.DOI:10.13203/j.whugis20210377

# 卷积神经网络滑窗识别结合语义推理的城市功能区分类

王严<sup>1</sup> 刘万军<sup>2</sup> 谭亚丽<sup>1</sup> 李玉<sup>1</sup>

1 辽宁工程技术大学测绘与地理科学学院,辽宁 阜新,123000

2 辽宁工程技术大学软件学院,辽宁 葫芦岛,125000

**摘要:**目前基于遥感图像的城市功能区分类方法通常采用光谱特征解译、兴趣点数据辅助、评价策略优化等方式,依赖大量人工操作,并借助遥感图像外的其他信息源。为了解决以上问题,提出利用卷积神经网络进行滑窗识别,提取图像语义标签,结合语义推理机制实现城市功能区分类的滑窗-推理方法。首先,建立两级城市功能区分类,以二级城市功能区为标识标注训练样本,并以此训练卷积神经网络作为识别器;然后,设计有重叠的滑窗识别模式,使用识别器辨识滑窗区域内图像块的二级城市功能区类型;最后,提出一个带权重的打分机制,作为语义推理方式,语义推理对象为全部识别结果,确定各图像块的一级城市功能区类型,实现遥感图像城市功能区分类。实验使用模拟图像和高分辨率遥感图像,两种图像的总分类精度分别可达94.50%、92.04%。滑窗-推理方法旨在通过语义推理处理滑窗识别产生的多语义标签,根据多语义标签确定对象的真实城市功能区。实验结果表明,所提方法无需辅助信息,直接利用遥感图像进行城市功能区分类是可行和有效的。

**关键词:**城市功能区;图像分类;卷积神经网络;滑窗识别;语义推理;投票机制

中图分类号:P237

文献标识码:A

收稿日期:2022-02-27

DOI:10.13203/j.whugis20210377

文章编号:1671-8860(2023)06-0950-10

## Classification of Urban Functional Areas by Convolution Neural Network Recognition Combined with Sliding Window and Semantic Reasoning

WANG Yan<sup>1</sup> LIU Wanjun<sup>2</sup> TAN Yali<sup>1</sup> LI Yu<sup>1</sup>

1 School of Geomatics, Liaoning Technical University, Fuxin 123000, China

2 School of Software, Liaoning Technical University, Huludao 125000, China

**Abstract:** Objectives: Although remote sensing image is one of the main data sources for the classification of urban functional areas, it is rare to use remote sensing images to classify urban functional areas and extract their attribute information. At present, the classification methods of urban functional areas based on remote sensing images usually need manual interpretation, point of interest data assistance, information questionnaire survey and so on. This kind of method not only needs a lot of manual operation, but also needs other external information sources except remote sensing images. Methods: In order to solve the above problems, a sliding window-reasoning method for urban functional area classification is proposed by using convolutional neural network (CNN) to identify sliding window, extract image semantic tags, and combine semantic reasoning mechanism, which can quickly realize urban functional area classification only by remote sensing images. First, a two-level classification table of urban functional areas is established, and the training samples are marked with the second-level urban functional areas, and the training CNN is used as

基金项目:国家自然科学基金(41801368);辽宁省教育厅重点项目(LJ2020ZD003);辽宁省教育厅科研经费项目(LJKZ0350)。

第一作者:王严,博士生,主要从事遥感图像信息提取研究。wangyan\_mapping@163.com

通讯作者:李玉,博士,教授。liyu@lnut.edu.cn

the recognizer. Second, using the designed overlapping sliding window recognition pattern, the trained recognizer is used to identify the types of features in the sliding window area, and to determine the type of urban functional area. Finally, a weighted scoring mechanism is designed as the implementation of semantic reasoning, and the semantic reasoning objects are all the recognition results. And, the type of urban functional areas of each region is determined, and the urban functional areas of large-scale remote sensing images are classified. **Results:** Using simulated images and high-resolution remote sensing image experiments, the total classification accuracy of simulated image experiments based on confusion matrix can reach 94.50%, and the total classification accuracy of real remote sensing image experiments based on confusion matrix can reach 92.04%. **Conclusions:** The purpose of the sliding window-reasoning method is to deal with the multi-semantic label results produced by the sliding window recognition through the semantic reasoning method, and to determine the real urban functional area results of the identified objects through multiple semantic tags. The results show that the method of CNN sliding window recognition combined with semantic reasoning is feasible and effective to classify urban functional areas directly using large-scale remote sensing images without the assistance of other information.

**Key words:** urban functional area; image classification; convolutional neural network; sliding window recognition; semantic reasoning; voting mechanism

中国城镇化规模不断扩大,城市发生着巨大的变化,在城市规划、土地利用政策制定、城市建设等方面产生诸多问题。城市土地利用类型结合城市中居民的实际生活、生产需求,形成各类城市功能区,城市功能区自动精准检测可为这些问题的解决提供基础数据<sup>[1-2]</sup>。以往市政部门往往关心土地利用现状的检测和分类,而忽略其使用功能的变化。从功能区的角度规划城市,不仅是城市建设的重要问题,也是方便城市居民生产、生活的重大民生问题<sup>[3-4]</sup>。随着遥感技术高速发展,其图像分辨率不断提高,利用高分遥感图像实现城市功能区的检测成为可能<sup>[5-6]</sup>。

目前基于遥感图像的城市功能区分类方法普遍利用光谱特征、兴趣点(point of interest, POI)、评价策略等辅助信息实现<sup>[7-9]</sup>。文献[4]基于POI开放性数据,利用核密度、空间主成分等工具,分析城市服务的布局特点,为城乡规划应用提供新方法;文献[10-11]采用重尾打断分类法和核密度聚类法对POI进行统计并建模,进行城市功能区识别;文献[12-13]基于时空活动数据研究城市功能结构分类。上述方法均需遥感图像以外数据才可进行。

近年来,卷积神经网络(convolutional neural network, CNN)受到学者的青睐,并出现许多利用CNN进行遥感图像城市功能区分类的方法<sup>[14-15]</sup>。文献[14]采用超像素划分与CNN结合的方式实现城市功能区分类,其精度受分割方法制约,不规则的超像素块重采样后作为CNN的输入也对精度有一定的影响。LeNet-5提出卷

积-池化模型<sup>[16]</sup>,VGG Net 和 Google Net 将 Alex Net<sup>[17]</sup>改进为更深的网络模型,在精度上均有较大提高<sup>[18-19]</sup>。VGG Net 以多个  $3 \times 3$  的卷积核来代替较大的卷积核,从而加快图像的运算速度,并提取更抽象的特征。随着网络加深,得到抽象的语义特征更容易描述图像。Google Net 采用网中网(network in network, NIN)的网络设计思路和 Inception 结构,更好地利用了网络中的计算资源,并在不增加计算负荷的情况下,增加网络的宽度和深度。随着计算机性能的提升,学者们提出图像上采样技术,出现了以全卷积神经网络(fully convolutional network, FCN)为代表的像素级深度学习图像分类算法<sup>[20]</sup>,但其样本标注过程困难,考虑到城市功能区语义复杂,利用FCN训练极易出现过拟合,不利于泛化。

CNN 只可处理固定尺度的图像,不适合直接用于大尺度复杂图像的识别。因此,定义一种与目标检测相似的滑窗方式应对此问题,即运用一固定尺度的视窗,自图像左上角开始依一定步长向右滑动,遍历一行后向下滑动同样的步长<sup>[21]</sup>。如此,复杂图像转化为大量相对简单、大小相同的小图像,继而利用已训练完备的 CNN 网络模型进行识别<sup>[17]</sup>。面对杂乱的识别结果,使用语义推理的方法进行处理,完成整幅图像的城市功能区分类工作。滑窗-推理方法将多个地物视为一个整体进行研究,不仅可用于城市功能区分类,还可探究大尺度图像复杂语义对象的快速分类方法。采用重叠滑窗方法提取图像中的语义,并对语义进行挖掘,是大数据的一种新处理模式。

# 1 CNN滑窗识别及语义推理方法

## 1.1 识别城市功能区的CNN模型

本文采用的CNN模型为改进的Alex Net模型,进行多次网络模型结构及参数调整,最终确定网络模型为一个10层神经网络模型,包括7个卷积层和3个池化层,如图1所示。

考虑输入图像尺度问题,加大前两个卷积组

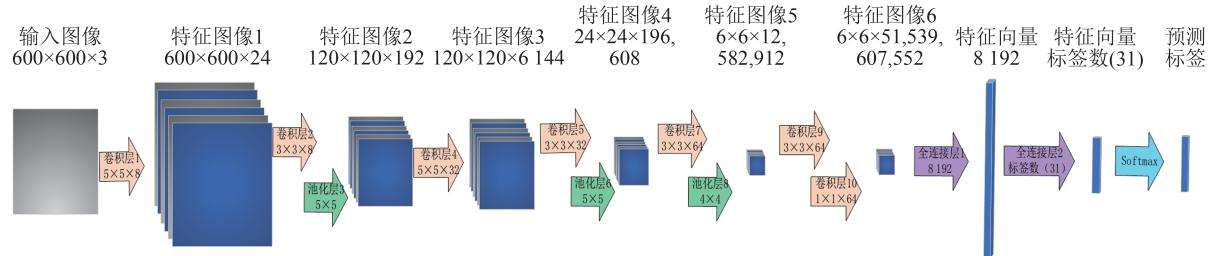


图1 网络模型

Fig. 1 Network Structure

在网络模型中通过减少卷积核数量来降低卷积模块的参数量,选用ReLU函数作为激活函数<sup>[23]</sup>。同时考虑到处理的图像尺度为600×600像素,因此加大池化层宽度可以达到快速降低特征维度的目的,网络特征图的最终维度为6×6像素,减少了全连接层的参数量并加速训练过程。另外,加宽网络模型第一层全连接层的宽度,最后一层的全连接层神经元数量改为标签数量,并使用Softmax函数作为输出层函数,训练改进后的网络模型作为遥感图像的场景识别器。

## 1.2 样本填充及训练

本文以航空影像数据集(aerial image data, AID)<sup>[24]</sup>作为基础数据集,该数据集共包含30个场景类别。在实验中发现,数据集缺少海洋样本,样本完整性不足。此外数据集中部分图像分辨率较低,无法满足高分辨率图像的场景分类工作。因此在高分辨率遥感图片上采样,通过人工判读并对比地图数据,确定每个样本场景类别,并填充到AID数据集中。将一张遥感图像描述为一个张量,其高度为所处理图像的通道数。训练中,每幅图像被认为只包含一种场景,每种场景表达为二级分类类别。

在填充后的数据集中随机选取80%样本作为训练集,其余作为验证集。训练改进的CNN,训练后模型识别精度超过80%。利用训练完备的CNN作为识别器分类待处理图片,分类结果为某一城市场景,即二级城市功能区类别,如图2所示,其识别结果为密集居民区。根据Softmax

(即卷积层1、2及卷积层4、5)的感受野,设置两个卷积核尺度分别为5×5和3×3,并在每个卷积组后进行尺度为5×5的池化操作,对图像进行下采样。图像尺度衰减后采用3×3大小的卷积核继续提取深层语义特征,并在全连接层前加入1×1卷积,对提取到的特征图进行滤波处理。经两个全连接层及Softmax函数处理<sup>[22]</sup>,得到预测标签。



图2 二级分类识别结果

Fig. 2 Identification Results of Secondary Classification

## 1.3 功能区滑窗识别器

### 1.3.1 滑窗识别

采用与目标检测相似的滑窗方式,利用已训练的CNN网络模型对功能区进行滑窗识别。滑窗方式分为重叠与无重叠两种,若每次滑窗没有重叠,原图可视为被分割若干块。一幅图像中可能存在多种城市功能区场景,但Softmax处理后,只会返回其中占图像主体的城市功能区类型标签。在识别过程中,图像主体识别较为准确,一般认为图像中心区域为图像主体,因此无重叠滑窗识别出的功能区在各块边缘处识别结果可靠

性低,无重叠滑窗的识别结果过于武断。另外,识别块尺度较大时,采用无重叠滑窗方式得到的识别结果必然十分粗糙,所以设定固定步长值的重叠滑窗方式更加合理。

给定一幅图像  $I = \{I_{ij}(R_{ij}, G_{ij}, B_{ij}) ; i=1, 2, \dots, z_1, j=1, 2, \dots, z_2\}$ , 其中,  $I_{ij}$  为像素索引;  $R_{ij}, G_{ij}, B_{ij}$  分别为此像素的红、绿、蓝三通道的值;  $i$  和  $j$  分别为像素横、纵坐标索引;  $z_1$  和  $z_2$  分别为图像横、纵像素数。图 3 为滑窗示意图。首先用大小为  $L \times L$  的视窗,从图像左上角第一个像素开始依步长  $P$  向右滑动,直到滑窗右边缘至图像右边缘;然后滑窗回到索引  $i$  为 1 位置,向下滑动步长  $P$ ,之后向右滑动;依此类推,直至完成整幅影像。滑窗对应像素块为  $K = \{K_{uv} = I_{uv}; u=1, 2, \dots, k_1, v=1, 2, \dots, k_2\}$ , 其中  $K_{uv}$  为滑窗索引;  $u$  和  $v$  为其横、纵坐标索引;  $k_1, k_2$  为横、纵滑窗最大滑动次数,  $I_{uv} = \{I_{ij}; i \in [u, u+P), j \in [v, v+P)\}$ 。将每个像素块输入识别器中进行识别,记录其识别结果  $S = \{S_{uv}(u, v); u = 1, 2, \dots, k_1, v = 1, 2, \dots, k_2\}$ , 并作为对应像素块的语义标签。

由图 3 中可以看出,当滑窗操作完成时,由于滑窗重叠,滑窗堆叠形成许多正方形图像区域块,其边长为滑动步长  $P$ 。定义滑窗范围内的图像块为识别块,在一定数量的识别块堆叠后,重叠区域为正方形,此区域有且只有一个,定义为处理块。形成一个处理块的识别块数量由滑窗大小  $L$  与步长  $P$  确定,其值为  $L$  与  $P$  比值的平方。同一处理块可在多个识别块中,但在不同的识别块中可能得到不同的语义标签。

同一处理块的识别差异如图 4 所示,同一个处理块在两个识别块中得到不同的识别结果,处理块将获得不同的语义标签。需要语义推理解决复杂语义问题,即根据包含同一处理块的所有识别块(堆叠出此处理块的所有识别块)的识别结果,确定此处理块的功能区类别。模型中滑窗大小  $L$  与识别器的输入相同,而步长  $P$  应根据图像空间分辨率确定。处理块作为模型的基本处理单位,其大小应该与功能区最小面积相当。由于城市功能区主要为居民区和商业区,顾及其建筑的占地面积,认为处理块覆盖真实地面范围应大于  $900 \text{ m}^2$ 。另外,选择  $P$  值时应考虑  $L$  与  $P$  的比值是否为整数,综合以上因素确定  $L$  和  $P$  的值。

### 1.3.2 图像填充处理及识别过程

滑窗识别时,处理块识别次数如图 5 所示。

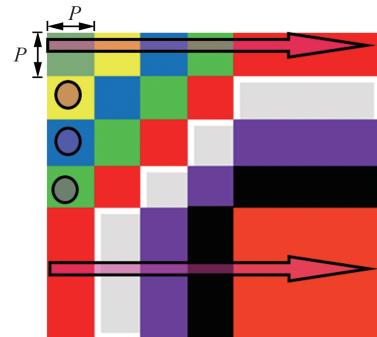


图 3 滑窗示意图

Fig. 3 Diagram of Sliding Window



图 4 识别差异示意图

Fig. 4 Diagram of Identify Difference in Results

1	2	3	3	3	3	2	1
2	4	6	6	6	6	4	2
3	6	9	9	9	9	6	3
3	6	9	9	9	9	6	3
3	6	9	9	9	9	6	3
3	6	9	9	9	9	6	3
2	4	6	6	6	6	4	2
1	2	3	3	3	3	2	1

图 5 处理块识别次数

Fig. 5 Number of Processing Blocks

由图 5 可以看出,图像边缘区域识别次数少,结果可靠性低且无法形成处理块;图像中间区域的识别次数相同,为使识别区域位于图像中间区域,通常需填充图像。由于 CNN 识别不关心特征位置,只关心有无特征,采用镜像方式填充边缘,既保证原图区域为识别区域,又保证识别结果的可靠性。在处理超大尺度图像时,仅需选定感兴趣区,不必填充处理。

由于 CNN 识别对象必须等大,而所有滑窗均须在处理图像内,无法保证所有图像可以满足以上条件,故需要第二次填充。根据图 3 可知,第一次填充值可由识别块横、纵像素数  $H$  以及步长  $P$  确定,滑窗中所有识别块均为正方形,填充值  $T_1$  的计算式为:

$$T_1 = H - P \quad (1)$$

得到第一次填充后图像的横、纵像素数  $X_1$ 、 $Y_1$ 之后,对图像进行镜像填充,第二次填充值  $T_{2X}$  和  $T_{2Y}$  的计算式为:

$$\begin{cases} T_{2X} = f_{ceil}(X_1/P) \times P - X_1 \\ T_{2Y} = f_{ceil}(Y_1/P) \times P - Y_1 \end{cases} \quad (2)$$

式中,  $f_{ceil}()$  为向上取整函数。

填充后依前文方法进行滑窗识别,根据二级功能区标签确定一级功能区类别,储存其一级功能区类别为识别结果。将一级功能区类别赋值一个数字作为识别块的标签,按照识别块位置储存至矩阵,得到识别结果矩阵。由于图像经过填充,滑窗间有重叠,无法将识别结果作为原图功能区分类结果。即使可以,得到的功能区分类结果必然粗糙,结果图的大小也与原图不符,所以必须引入处理块,使处理块拼接结果与原图尺度相同,分类更加精细。根据识别块的标签,通过语义推理确定每个处理块的功能区属性,完成功能区的分类。

### 1.3.3 语义推理

根据 CNN 识别中心区域更准确的特点,设计一个打分系统对每个处理块进行语义推理,具体推理过程如下:

1) 以识别结果矩阵中包含同一处理块的全部识别块对应区域为打分区域,将打分区域内的子矩阵取出并旋转 180° 作为打分矩阵,打分矩阵的每个元素(标签)在矩阵中的位置与该处理块在对应识别块中位置相同。

2) 判断标签在打分矩阵中的位置,根据位置获得不同分数。当其处于中心位置时,标签对应的类得分更高。假设共有 3 类功能区,创建记分器  $F_1, F_2, F_3$  与其一一对应。设置识别块处在最外层时分为 1, 即该位置得分时, 对应功能区的记分器分值 +1。为了凸显中间位置重要性, 内层每一个元素的得分大于外层元素得分之和, 设置当前层每一元素得分是外层所有元素之和的  $5/4$ 。考虑识别块和处理块均为正方形, 分数计算式为:

$$F_m = \prod_{n=1}^R \max \left[ 1, (d - 2n + 1) \times 5 \times \frac{\max(0, m - n)}{\max(1, m - n)} \right] \quad (3)$$

式中,  $F_m$  为当前层元素得分; 以最外层元素组成的层为第一层,  $n, m, R$  分别为层数索引、当前元素所处层数、总层数;  $d$  表示识别块与处理块边长的比值,  $d = L/P$ 。

3) 搜索 3 种功能区识别结果在各个层的数量,统计记分总和。比较 3 个记分器得分,将此处理块标签设置为得分最高的记分器对应的功能区类别,依次确定所有处理块标签。

本文方法的实验流程如图 6 所示。假设一个处理块可由  $6 \times 6$  个识别块决定, 该处理块的完整识别过程如图 7 所示。其中 36 个识别块重叠出 1 个处理块, 第一个和最后一个识别块的位置记为初始位置和最终位置。由图 7 可以看出, 推理结果中处理块的对应位置与原图位置一致。

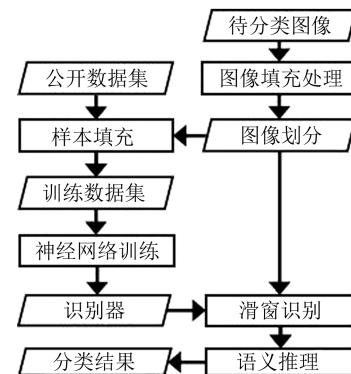


Fig. 6 Experimental Flowchart of the Proposed Method

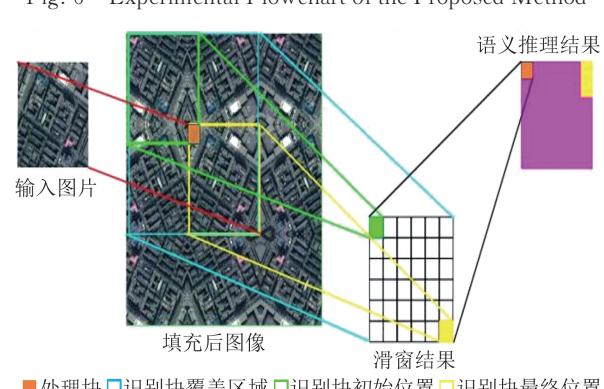


Fig. 7 Identification Process

## 2 实验与结果分析

### 2.1 模拟图像实验

为验证算法有效性, 建立彩色数据集进行模拟图像实验。模拟图像和实验结果如图 8 所示, 其中模拟图 1 的尺寸为  $8960 \times 8960$ , 以 5 种颜色代表 5 种功能区, 即非功能区、居民区、商业区、工业区和农业区。为使模拟图像接近真实遥感图像, 模拟图 2 中几何图形以无规律方式排列, 并加入大量噪声以增加图像复杂度, 图像尺寸为  $6000 \times 6000$ 。两张模拟图像的实验超参数相同, 设置滑窗大小为  $600 \times 600$ , 滑动步长为 100, 采用改进的网络模型训练, 得到识别器, 获得分类结果。

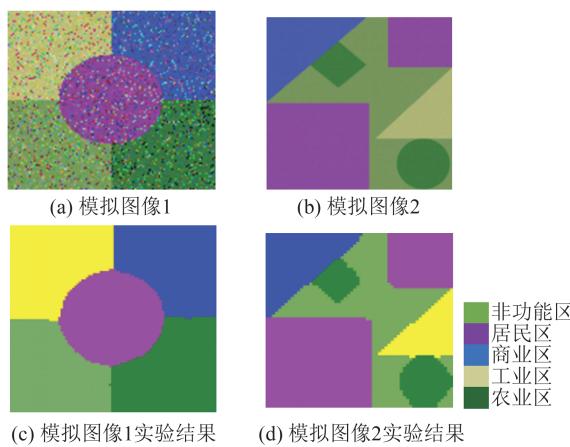


图8 模拟图像和实验结果

Fig. 8 Simulated Image and Experiment Result

为模拟真实遥感图像繁多的地物种类和复杂的地物结构,在模拟图像中加入大量噪声,不仅验证算法抗噪性,还提高图像的复杂程度。但即使加入多种噪声,也无法完全模拟真实遥感图像,对于滑窗-推理方法而言,模拟图像识别工作更加容易。由图8(c)和图8(d)可以看出,在同一颜色区域内,几乎没有识别错误;结合图8(a)可以证明模型在分类工作上的有效性和可泛化性;由图8(d)可见,由于滑窗方法步长的限制,实验结果中出现锯齿状边界,这是由于语义推理的投票采用高分获胜原则,致使几何图形边角处出现缺损或突出区域。

模拟图像2的分类精度见表1,其总分类精度为94.50%,Kappa系数为0.9258。由表1可知,商业区、居民区、工业区的产品精度和用户精度均较高;面积最小的农业区产品精度较低,用户精度较高;而非功能区分布最分散,其产品精度

和用户精度均在90%左右,相对其余4类精度较低。分析其原因,商业区、居民区、工业区比较集中,在投票过程中相同类别地物集中将获得得分优势,因此对于同一功能区集中的区域滑窗-推理方法将得到更好的结果。非功能区分布分散,并与每种区域都有接触,其分类精度不单受到语义推理结果影响,还会在滑窗识别过程中受到其他区域的干扰。

表1 模拟图像2的分类精度

Tab. 1 Classification Accuracy of Simulated Image 2

精度指标	地物类别				
	农业区	商业区	居民区	工业区	非功能区
产品精度	0.8357	0.9755	0.9966	0.9294	0.9093
用户精度	0.9625	0.9927	0.9467	0.9815	0.9025

由模拟图像实验可知,在识别结果准确的前提下,滑窗-推理方法可以有效地进行分类。分类精度主要受滑窗步长和识别器精度影响,此外,受推理方式影响,还会出现规则几何图形的边角处缺损、边缘位置呈锯齿状等问题。

## 2.2 真实图像实验

选用两幅真实遥感图像进行实验,实验结果如图9所示。其中真实图像1为WorldView3卫星获取的巴西某地区RGB城市遥感图像,空间分辨率为0.31 m,图像大小为8 837×8 823像素;真实图像2为中国辽宁省沈阳市某区域的高分二号(GF-2)卫星数据,利用多光谱数据和全色数据合成空间分辨率为0.8 m的高分辨率RGB图像,图像大小为10 000×10 000像素。依城市情况,设置一、二级城市功能区分类,见表2。

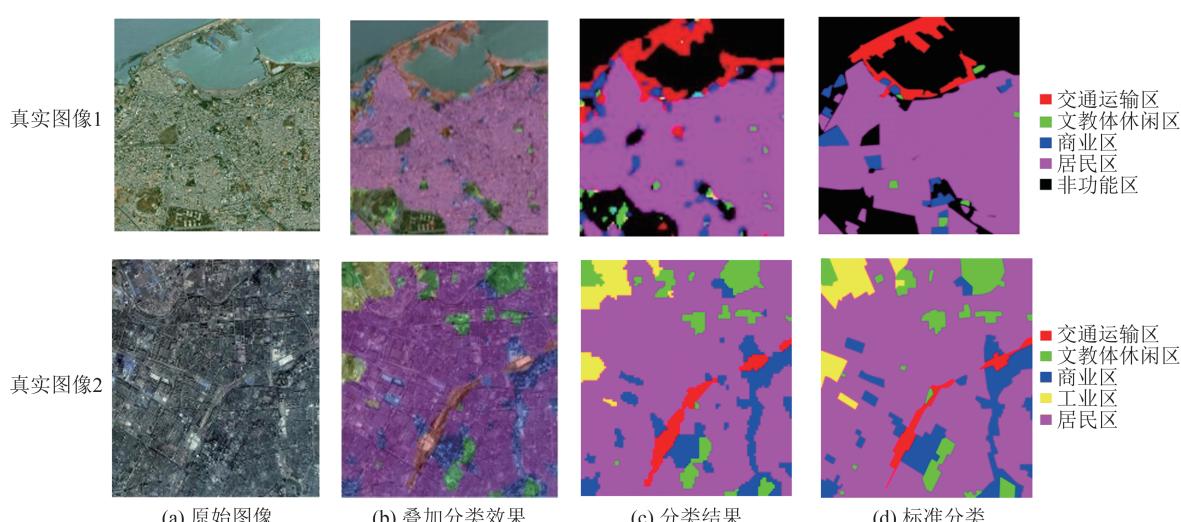


图9 遥感图像和实验结果

Fig. 9 Remote Images and Experiment Results

表2 功能区分类语义表

Tab. 2 Classification Semantic of Urban Functional Areas

一级分类	二级分类
居住区	城市社区(稀疏、中等住宅)、农村(密集住宅)等
文教体休闲区	学校、行政机关、教堂、公园、体育场、旅游区等
商业区	商业中心、商场等
工业区	工厂、仓储等
交通运输区	机场、港口、火车站等
农业区	农田、农业大棚等
非功能区	裸地、海滩、沙漠、森林、草地、山、停车场、池塘、河流等

实验程序采用 Matlab R2018b 编写,运行环境为 Windows10 64 位专业版操作系统,80 GB 内存,处理器为 Intel(R) Xeon(R) CPU E5-4650 图形图像工作站,搭载 NVIDIA Tesla K40c GPU 计算卡,训练改进的网络模型,得到识别器。顾及识别对象尺度,设置滑窗大小为  $600 \times 600$  像素,结合实验数据空间分辨率,参照§1.3 中模型识别块的设计,设计滑动步长为 100,即每个识别块大小为  $600 \times 600$ ,处理块大小为  $100 \times 100$ 。

在识别过程中,将各功能区赋予色彩标签,确定色彩表达系统,分类结果如图 9(c)所示;将实验图片进行人工标注,得到标准结果如图 9(d)所示;将分类结果与原图重叠,如图 9(b)所示。由图 9(b)可以看出,滑窗-推理方法可有效识别并分类城市功能区。将分类结果与人工标注结果对比,真实图像 1 的功能区分类准确度为 91.85%,Kappa 系数为 0.859 0;真实图像 2 的功能区分类准确度为 92.04%,Kappa 系数为 0.828 9。

两幅真实图像的分类精度见表 3、表 4,混淆矩阵如图 10 所示。由于城市的各功能区地物具有极高相似性,无法找到明确特征进行区别,使用语义特征也存在一定困难,因此本文建立多级语义表达,提取二级功能区语义,通过二级分类表完成功能区分类,此方式部分解决了功能区语义问题,但由于功能区的复杂性和多元性,多级语义不能根本解决该问题,尤其是在功能区交界处,其语义特征依然模糊,造成分类错误。另外,随着城市生态建设,功能区中包含道路、林地、草地等非功能区,当非功能区面积较大时,也影响功能区分类结果。

从各类精度分析,滑窗-推理方法在大面积功能区分类上有较好的表现,但对文教体休闲区、交通运输区等面积小的功能区,分类结果并不理想。究其原因,一方面,城市功能区语义复杂、场

景较多,识别问题导致分类错误;另一方面,小面积功能区空间上连续性不足,投票中处理块邻域对结果影响极大,小面积功能区必将受到此机制影响;此外,边界不准确也影响小区域的分类精度。在后续研究中,应改良语义推理机制,并设法平滑分类结果边缘,小区域的分类精度将有提高。综合分析,滑窗-推理方法可以完成城市功能区分类任务,但受语义识别能力影响严重,边缘问题也有待解决。

表3 真实图像 1 的分类精度

Tab. 3 Classification Accuracy of Experimental Image 1

精度指标	地物类别				
	交通运 输区	文教体 休闲区	非功能 区	商业区	居民区
产品精度	0.863 4	0.650 4	0.891 2	0.669 9	0.949 7
用户精度	0.659 6	0.483 4	0.977 6	0.478 5	0.944 6
交并比	0.597 2	0.290 4	0.863 6	0.387 2	0.933 1

表4 真实图像 2 的分类精度

Tab. 4 Classification Accuracy of Experimental Image 2

精度指标	地物类别				
	商业区	文教体 休闲区	交通运 输区	居民区	工业区
产品精度	0.823 3	0.871 5	0.646 7	0.954 7	0.888 0
用户精度	0.822 4	0.821 1	0.912 0	0.951 2	0.833 7
交并比	0.699 0	0.732 4	0.608 7	0.910 2	0.754 4

### 2.3 对比实验与分析

在高分遥感图像中,小尺度的图像块承载信息过少,城市功能区作为大量地物无固定规则的组合,只包含少量地物信息的图像块可能无法准确呈现城市功能区信息。滑窗-推理方法通过重叠滑窗识别城市功能区,其目的是利用大尺度识别块提高识别准确度,利用交叠出的处理块提高分类精度,通过语义推理确定处理块类别。

若采用与识别块等大的无重叠滑窗,可以保障识别准确度,难以得到精细的分类结果;欲得到精细程度相同的分类结果,可采用尺度与处理块大小相同的滑窗进行识别。针对以上两点,分别设计识别块尺度为  $L \times L$  和  $P \times P$  的两种无重叠滑窗实验,对比分析滑窗方式对实验的影响。以真实图像 1 为实验数据进行验证,对比实验滑窗尺度分别为  $600 \times 600$  和  $100 \times 100$ ,以相同的赋色方式对功能区进行标记,分类结果如图 11 所示。由于无重叠滑窗不会产生多标签结果,故无需进行语义推理,滑窗识别结果便是最终分类结果。

		预测分类				
		交通运输区	文教体休闲区	非功能区	商业区	居民区
真实分类	交通运输区	3663562	8810	120997	356550	93201
	文教体休闲区	16627	477987	18207	63467	158619
	非功能区	1078080	395170	31008693	888543	1422870
	商业区	126645	31221	98362	1660218	1561644
	居民区	669606	75612	473421	500622	55152917

		预测分类				
		商业区	文教体休闲区	交通运输区	居民区	工业区
真实分类	商业区	8715258	111173	299491	1471858	0
	文教体休闲区	212298	6431811	104804	1056935	27220
	交通运输区	70454	0	2325324	153934	0
	居民区	1587990	643324	865681	70318875	508364
	工业区	0	193853	200	653290	4277863

(a) 真实图像1

(b) 真实图像2

图 10 真实图像实验的混淆矩阵

Fig. 10 Confusion Matrix of Experimental Result

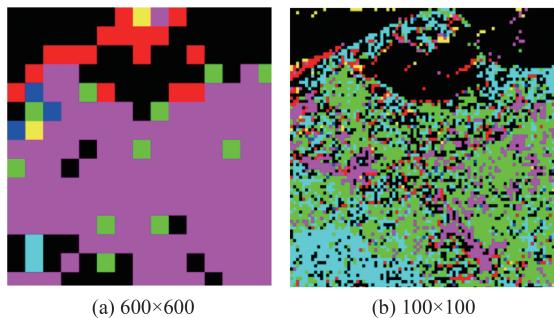


图 11 实验结果对比

Fig. 11 Comparision of Experiment Results

由图 11(a)可以看出,滑窗尺度与识别块尺度相同时,无重叠滑窗识别结果明显更加粗糙,小面积的功能区无法准确分类,边缘问题更加严重。将滑窗尺度降低到与处理块大小相同,验证是否可以得到与滑窗-推理方法相似结果。将 AID 数据集重新采样,并依§2.1 方法制作数据集,设计同深度 CNN 模型,得到精度相当的识别器,实验结果如图 11(b)所示。可以看出,图 11(b)的识别准确率极低,大量住宅区被识别为文教体休闲区,单建筑面较大的商业区几乎无法识别。分析其原因,实验图像空间分辨率为 0.31 m,100×100 像素的图像块覆盖真实地面约 900 m<sup>2</sup>,可传达地物信息较少。滑窗识别中,大面积地物无法完全覆盖,造成识别误差。此外,城市功能区识别不是单独地物识别,不同功能区存在大量相似地物,关注地物的同时需要顾及地物间的空间特征。滑窗尺度较小时,图像块传达地物信息过少,必然造成识别问题。

两次对比实验证明,大尺度无重叠滑窗识别造成的分类结果粗糙问题难以克服;小尺度无重叠滑窗分类结果较精细,但面对城市功能区识别问题,其识别能力不足。分析两次实验结果,无重叠滑窗难以进行城市功能区分类工作。

### 3 结语

为了解决网络模型只能进行图像级识别的问题,本文提出一种重叠滑窗处理与语义推理协同使用的方式。滑窗-推理方法实质是采用重叠滑窗对遥感图像进行方形滑窗并识别,得到语义标签后,利用语义推理方式解决重叠滑窗造成的多语义标签问题,以此完成城市功能区分类工作。从结果来看,滑窗-推理方法可以完成遥感图像城市功能区分类工作。

对比模拟图像实验和真实遥感图像实验不难发现,CNN 模型的识别能力对滑窗-推理模型的分类精度有巨大影响。真实遥感图像实验采用不同数据源、不同分辨率的两张高分遥感图像。在准确率和 Kappa 系数上均有相当的分类精度,足以证明其整体分类精度良好,具有一定的鲁棒性。以各功能区分类精度分析,模型在小面积独立功能区的分类问题上有明显不足。但目前研究中,神经网络模型的泛化性较差,难以满足任意分辨率、任意数据源的识别工作。如何优化识别器,是后续研究中的重点之一。

滑窗-推理方法现存在边缘不清晰,无法识别混合功能区等问题。对于边缘问题,目前存在以街区为对象的方法,可以引入街区的概念,通过提取路网的手段加强边缘,而如何更准确地识别城市功能区和混合功能区,是进一步优化的方向。将提取出的语义特征进行人为的二次挖掘,以实现对类似城市功能区这样的复杂对象的精准识别,也是值得思考的问题。滑窗-语义推理方式是以深度学习优异的识别能力为基础,通过多标签处理的方式完成分类工作,无论是滑窗还是多标签的处理形式和具体算法都有巨大改进空间。以推理方式为例,现采用的推理方式只考虑了 CNN 识别出的语义标签,语义推理方式也相

对简单,在今后的工作中,应考虑采用更好的推理方式,直接使用语义特征推理城市功能区类别。

## 参 考 文 献

- [1] Wu Haiping, Huang Shicun. Research on New Construction Land Information Extraction Based on Deep Learning: Innovation Exploration of the National Project of Land Use Monitoring via Remote Sensing[J]. *Remote Sensing for Land & Resources*, 2019, 31(4): 159–166. (吴海平, 黄世存. 基于深度学习的新增建设用地信息提取试验研究: 全国土地利用遥感监测工程创新探索[J]. 国土资源遥感, 2019, 31(4): 159–166.)
- [2] Chen Yimin, Li Xia. Applications and New Trends of Machine Learning in Urban Simulation Research [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12): 1884–1889. (陈逸敏, 黎夏. 机器学习在城市空间演化模拟中的应用与新趋势[J]. 武汉大学学报(信息科学版), 2020, 45(12): 1884–1889.)
- [3] Hao Wenxuan, Tong De, Liu Qing, et al. The Delimitation and Classified Planning and Management of Transformation Function Region: The Experience and Exploration of Urban Renewal in Shenzhen[J]. *Urban Development Studies*, 2015, 22(10): 42–48. (郝文璇, 全德, 刘青, 等. 改造功能区划定与分类规划管理: 来自深圳城市更新的经验和探讨[J]. 城市发展研究, 2015, 22(10): 42–48.)
- [4] Li Zhixue, Mo Wenbo, Zhou Songlin, et al. Urban Life Convenience Index Based on Spatial Principal Component Analysis: A Case Study of Changsha City[J]. *Journal of Geomatics*, 2021, 46(2): 110–115. (李志学, 莫文波, 周松林, 等. 空间主成分分析的城市生活便利度指数研究: 以长沙市为例[J]. 测绘地理信息, 2021, 46(2): 110–115.)
- [5] Zhang Huifang, Zhang Penglin, Chao Jian. Change Detection by Multi-scale Fuzzy Fusion on High Resolution Images [J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(2): 296–303. (张慧芳, 张鹏林, 晁剑. 使用多尺度模糊融合的高分影像变化检测[J]. 武汉大学学报(信息科学版), 2022, 47(2): 296–303.)
- [6] Xu Jinghai, Du Dongsheng, Li Zhijun, et al. A Real-time Dynamic Monitoring Method for Complex Building Applying Sensor Network and Reality 3D Model [J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(5): 630–639. (徐敬海, 杜东升, 李枝军, 等. 一种应用传感器网和实景三维模型的复杂建筑物实时动态监测方法[J]. 武汉大学学报(信息科学版), 2021, 46(5): 630–639.)
- [7] Zu Lihu, Yang Jing, Miao Shiyuan. Spatial-temporal Change of Land Use/Cover in Linxiang City Based on Remote Sensing Images [J]. *Journal of Geomatics*, 2021, 46(2): 30–35. (祖立辉, 杨静, 苗世源. 基于遥感影像的临湘市土地利用/覆盖时空变化研究[J]. 测绘地理信息, 2021, 46(2): 30–35.)
- [8] Qu Chang, Ren Yuhuan, Liu Yalan, et al. Functional Classification of Urban Buildings in High Resolution Remote Sensing Images Through POI-Assisted Analysis [J]. *Journal of Geo-Information Science*, 2017, 19(6): 831–837. (曲畅, 任玉环, 刘亚岚, 等. POI辅助下的高分辨率遥感影像城市建筑物功能分类研究[J]. 地球信息科学学报, 2017, 19(6): 831–837.)
- [9] Zheng Lifu, Shen Lingyun, Liu Huiping, et al. Quantitative Identification of Urban Mixed Functions Based on Multidimensional Data Obtained from Online Map [J]. *Journal of Geomatics*, 2022, 47(S1): 188–193. (郑力夫, 沈凌云, 刘慧平, 等. 基于在线地图数据的城市混合功能二三维划分方法研究[J]. 测绘地理信息, 2022, 47(S1): 188–193.)
- [10] Gu Yanyan, Jiao Limin, Dong Ting, et al. Spatial Distribution and Interaction Analysis of Urban Functional Areas Based on Multi-source Data [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(7): 1113–1121. (谷岩岩, 焦利民, 董婷, 等. 基于多源数据的城市功能区识别及相互作用分析[J]. 武汉大学学报(信息科学版), 2018, 43(7): 1113–1121.)
- [11] Huang Yimin, Shao Shiwei, Lei Yingzhe, et al. Identification of Urban Functional Zones Using Network Kernel Density Estimation and Kriging Interpolation [J]. *Journal of Geomatics*, 2019, 44(4): 14–18. (黄怡敏, 邵世维, 雷英哲, 等. 运用网络核密度估计与克里格插值识别城市功能区[J]. 测绘地理信息, 2019, 44(4): 14–18.)
- [12] Peng Zhenghong, Sun Zhihao, Cheng Qing, et al. Urban Land Use Function Recognition Method Using Sequential Mobile Phone Data [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(9): 1399–1407. (彭正洪, 孙志豪, 程青, 等. 利用时序手机通话数据识别城市用地功能[J]. 武汉大学学报(信息科学版), 2018, 43(9): 1399–1407.)
- [13] Yao Yao, Zhang Yatao, Guan Qingfeng, et al. Sensing Multi-level Urban Functional Structures by Using Time Series Taxi Trajectory Data [J]. *Geomatics and Information Science of Wuhan University*

- ty, 2019, 44(6): 875–884. (姚尧, 张亚涛, 关庆锋, 等. 使用时序出租车轨迹识别多层次城市功能结构[J]. 武汉大学学报(信息科学版), 2019, 44(6): 875–884.)
- [14] Zhou W, Ming D P, Lü X W, et al. SO-CNN Based Urban Functional Zone Fine Division with VHR Remote Sensing Image[J]. *Remote Sensing of Environment*, 2020, 236: 111458.
- [15] Yu J, Zeng P, Yu Y Y, et al. A Combined Convolutional Neural Network for Urban Land-Use Classification with GIS Data[J]. *Remote Sensing*, 2022, 14(5): 1128.
- [16] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based Learning Applied to Document Recognition [J]. *Proceedings of the IEEE*, 1998, 86 (11) : 2278–2324.
- [17] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Communications of the ACM*, 2016, 60 (6): 84–90.
- [18] Kim J, Lee J K, Lee K M. Accurate Image Super-resolution Using very Deep Convolutional Networks [C]// IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016.
- [19] Szegedy C, Liu W, Jia Y Q, et al. Going Deeper with Convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015.
- [20] Tian T, Chu Z, Hu Q, et al. Class-Wise Fully Convolutional Network for Semantic Segmentation of Remote Sensing Images [J]. *Remote Sensing*, 2021, 13(10): 3211.
- [21] Hu Min, Zhang Keke, Wang Xiaohua, et al. Video Facial Expression Recognition Combined with Sliding Window Dynamic Time Warping and CNN [J]. *Journal of Image and Graphics*, 2018, 23(8): 1144–1153. (胡敏, 张柯柯, 王晓华, 等. 结合滑动窗口动态时间规整和CNN的视频人脸表情识别[J]. 中国图象图形学报, 2018, 23(8): 1144–1153.)
- [22] Gimpel K, Smith N A. Softmax-margin CRFS: Training Log-linear Models with Cost Functions [C]// Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, USA, 2010.
- [23] Eckle K, Schmidt-Hieber J. A Comparison of Deep Networks with ReLU Activation Function and Linear Spline-type Methods [J]. *Neural Networks: the Official Journal of the International Neural Network Society*, 2019, 110: 232–242.
- [24] Xia G S, Hu J W, Hu F, et al. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3965–3981.