

doi: 10.11835/j.issn.1000-582X.2022.11.010

# 基于 Segnet 网络和迁移学习的全景街区影像变化检测

余晓娜<sup>1</sup>, 黄 亮<sup>1,2</sup>, 陈朋弟<sup>2</sup>

(1. 昆明理工大学 国土资源工程学院, 昆明 650093; 2. 云南省高校高原山区空间信息  
测绘技术应用工程研究中心, 昆明 650093)

**摘要:**针对采用传统方法难以提高全景街区影像变化检测精度的问题,提出一种基于 Segnet 网络和迁移学习的全景街区影像变化检测方法。首先对数据集“TSUNAMI”进行预训练并对训练集进行分类归并;然后采用 Segnet 网络对全景街区影像进行语义分割,最后对语义分割结果进行差值运算,得到变化差异图并进行精度评价。实验选取两组全景街区影像,采用最大似然法、支持向量机方法(SVM, support vector machine)以及提出方法对这 2 组数据进行对比实验,第一组得到的精度分别为 65.1%、72.1%和 81.4%;第二组得到的精度分别为 66.5%、70.6%、82.2%。实验结果表明提出的方法具有更高的变化检测精度,可为城市违章排查、灾后重建等提供技术支撑。

**关键词:** Segnet 网络; 迁移学习; 全景街区影像; 变化检测; 支持向量机

中图分类号: P237

文献标志码: A

文章编号: 1000-582X(2022)11-100-08

## Complex street scene change detection based on segnet network and migration learning

YU Xiaona<sup>1</sup>, HUANG Liang<sup>1,2</sup>, CHEN Pengdi<sup>2</sup>

(1. Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, P. R. China; 2. Surveying and Mapping Geo-Informatics Technology Research Center on Plateau Mountains of Yunnan Higher Education, Kunming 650093, P. R. China)

**Abstract:** The use of multi-temporal panoramic block images is of great significance for monitoring urban development and assisting government decision-making. However, due to the influence of solar rays, ground spectrum and shooting angle during the process of collecting data, it is difficult to obtain high precision by traditional methods. Complex neighborhood changes information. To this end, this paper proposes a method for detecting image change in panoramic blocks based on Segnet and migration learning. Firstly, the data set “TSUNAMI” is pre-trained and the training set is classified and merged. Then, the Segnet network is used to semantically segment the panoramic block image, and the semantic segmentation result is subjected to difference calculation to obtain the change result map and evaluate the accuracy. Experiments were carried out to select two groups of panoramic block images. The maximum

收稿日期: 2020-01-22 网络出版日期: 2021-06-15

基金项目: 国家自然科学基金资助项目(41961039); 云南省应用基础研究计划面上项目(2018FB078); 自然资源部地球观测与时空信息科学重点实验室经费资助项目(201911)。

Supported by National Natural Science Foundation of China (41961039), Applied Basic Research Programs of Science and Technology Department of Yunnan Province (2018FB078), and Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology of Ministry of Natural Resources(201911) .

作者简介: 余晓娜(1995—), 女, 硕士, 主要从事遥感影像变化检测方向研究, (E-mail)1304953659@qq.com。

likelihood method, the support vector machine method and the method proposed in this paper were used to compare the two groups of data. The accuracy of the first group was 65.1%, 72.1% and 81.4%, respectively. The accuracy of the second group was 66.5%, 70.6%, and 82.2%, respectively. The experimental results show that the proposed method has higher detection accuracy and can provide technical support for urban violation investigation, post-disaster reconstruction, and ancient cultural relics restoration.

**Keywords:** Segnet network; migration learning; panoramic block image; change detection; support vector machines

街区作为城市建筑物的主要组成部分,精确、实时地对道路街区进行变化提取,对于城市规划和土地利用调查具有十分重要的作用。例如,在城市发展中所出现的居民区改造、工业区向郊区迁移、商业区整改等。全景街区影像研究通过不同时期影像的获取,可以对城市街道的违建、违停进行排查,保证一个规整有序的城市环境。对全景街区影像进行变化检测,不仅能辅助城市的违章排查、灾后评估,还能检测出城市土地利用的变化情况,为政府在城市建设方面提供科学合理的理论参考。全景街区影像是指通过车载相机所采集到的 360°街区全视场影像<sup>[1]</sup>。但是由于全景影像所包含的地物复杂,不仅包括建筑物上的玻璃幕墙,还包括一些废墟、电线杆、交通灯等,所以传统的方法不适用于全景街区影像的变化检测。

目前,有很少一部分学者对全景街区影像的变化检测邻域进行研究。主要原因是对全景街区的变化检测存在“语义鸿沟”问题<sup>[2-3]</sup>。轩永仓等基于 FCN 网络模型实现图像像素级的预测,为复杂场景图像的语义分割做了良好铺垫<sup>[4]</sup>。Wu C 等运用贝叶斯理论和基于规则的方法,提高了场景变化检测准确率<sup>[5]</sup>。Arabi Mohammed El Amin 等提出了一种基于 CNN 特征的卫星图像变化检测方法,并得到较高精度<sup>[6]</sup>。刘文涛等通过级联式 FCN 和空洞卷积的方法,实现建筑物屋顶的精确分割和提取<sup>[7]</sup>。魏杨等基于深度学习识别出初步候选区,其次通过 Fast R-CNN 网络框架,实现农作物虫害的精准识别<sup>[8]</sup>。邓国徽提出的基于改进的 FCN 网络模型算法准确地识别出施工场地<sup>[9]</sup>。Nicolas 等通过引入多核卷积,并基于 SegNet 框架执行准确的语义分割,最终实现全景影像的准确标记<sup>[10]</sup>。虽然深度学习在遥感领域的应用比较广泛,也得到了比较理想的效果,但目前国内外对于全景街区影像的变化检测研究则相对较少。

笔者采用全景街区影像作为实验数据,并结合深度学习和迁移学习的思想,提出了基于 Segnet 网络的全景街区影像变化检测。相比传统方法,采用 SegNet 进行语义分割,可以更好地区分出不同地物目标,更加准确地进行地物提取。采用迁移学习的思想,可以大大缩短训练时间,提高实验精度,还为全景街区影像的变化检测研究提供理论参考。该方法在语义层次上检测到区域的变化情况,对现实中的街道违建排查、土地利用情况以及城市规划方面有重要意义。

## 1 全景街区影像变化检测

笔者研究了基于 Segnet 网络和迁移学习的全景街区影像变化检测。首先,对数据集“TSUNAMI”做预训练;其次,对训练集进行分类归并,分类归并的主要依据是地物的光谱、纹理等特征;然后,通过 Segnet 网络对实验数据进行语义分割,得到语义分割结果图;最后,对 2 幅语义分割结果图进行差值运算,得到最终的变化结果图,并对实验结果进行精度评价。与传统的方法相比,先对数据做训练,对训练模型做信息的分类归并,再进行语义分割,较传统方法得到较高精度。采用 Segnet 网络做语义分割使变化信息提取更加准确,并对类和类之间的区分也更加明显。另一方面,Segnet 网络和迁移学习相结合,大大减少了实验的工作量,缩短实验时间,提高实验结果的精度。该方法不仅适用于全景街区影像的变化检测,也适用于道路、建筑物、交通标志等信息的提取,有较大发展潜力,其技术路线图如图 1 所示。

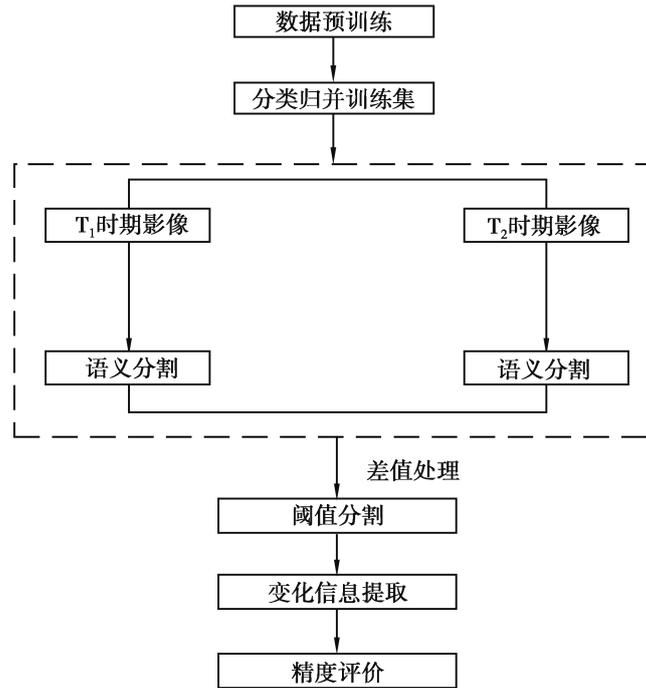


图 1 技术路线图

Fig. 1 Technical roadmap

### 1.1 Segnet 神经网络

Hinton G.E.在 2006 年第一次提出了深度学习的概念<sup>[11]</sup>。深度学习中常见的网络结构有 CNN、FCN、PSPNet、U-Net、Segnet 等,研究选取了目前应用比较成熟且广泛的 Segnet 网络模型作为实验模型。Segnet 网络的核心主要包括一个编码网络和一个与之对应的解码网络。Segnet 网络沿用了 FCN 图像语义分割的思想,并且该网络是基于像素级别的端到端网络架构。Segnet 沿用了 FCN 网络模型的思想,将 VGG16 中的全连接层去掉,将编码(encoder)信息和解码(decoder)信息直接连接,编码网络和解码网络作为整个网络结构的核心部分,其优点是保留了影像中大量有用的特征信息,使实验过程中需要训练的参数大大减少,缩减了实验数据的训练时间,最重要的是得到了相对较高精度的语义分割图像。

Segnet 神经网络结构如图 2 所示:Segnet 的网络结构主要包括卷积层(convolution)、归一化层(batch normalisation)、激活函数(ReLU)以及池化层(pooling)。

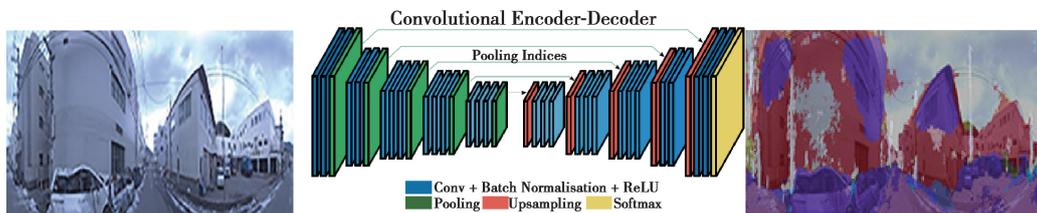


图 2 Segnet 神经网络结构图

Fig 2 Segnet neural network structure

从图 2 可以看出,Segnet 网络结构是一个对称的网络模型,网络的左边表示编码网络,右边表示解码网络。Segnet 网络的基本工作原理:在网络结构中,pooling 层与 upsampling 层主要进行影像分割,在整个网络结构中,特征地物的提取主要依靠位于左边的卷积层来完成,然而在提取的过程中,pooling 层的主要作用是使图片逐渐变小,这个过程被称为编码;右边的网络架构主要进行反卷积和 upsampling 操作,反卷积主要是使得图像的分类特征得以体现,而 upsampling 层主要是将分割后的影像恢复和原始输入图像一样的大

小,这个过程称为解码过程。通过编码网络不断提取特征,随之传输到相应的解码网络,对分割图像进行解码,最后再通过 softmax 分类器输出最终的语义分割结果图。

通过卷积运算,使影像中有用的特征信息更加突出,而忽略及削减影像中次要的信息,达到抑制噪声的目的。连接在卷积层之后的称为池化层,一般来说,池化层中特征图的个数和卷积层的特征图个数是保持一致的,两者之间是一一对应的关系。其中 Maximum pooling、Mean pooling、Random pooling 以及 Pyramid pooling 等是目前常用的池化方法<sup>[12-14]</sup>。

在 Segnet 的训练过程中,由于线性表达无法满足样本的多样性,以及实验过程中的复杂分类识别任务和训练数据集过大等原因,通常采用 ReLU 函数进行拟合。ReLU 激活函数是目前大多数卷积神经网络所采用的激活函数,ReLU 函数具有计算灵活、收敛快等特点,主要解决梯度下降的问题。其数学表达式为

$$f(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (1)$$

当输出信号大于 0 时,输出等于输入;当输出信号小于等于 0 时,输出等于 0。归一化层一般用于激活函数之前,最主要的作用是使学习的速度加快。

2010 年,Zeiler 等提出了反卷积的概念<sup>[15]</sup>。从网络结构上来说,反卷积层相当于一个上采样的过程,在训练过程中,通常由于全连接层的维度太大,训练的时候会出现参数增加而增大计算量的情况,为了解决这个问题,通常引入反卷积层。这样做的主要原因是因为反卷积层能把图像丢失的信息找回来,这样能最大限度的保留上下文信息,使得训练结果更加准确。

由于测试数据集较少,于是通过对图像的拉伸、旋转、平移等操作,增大数据集,把需要测试的数据集放进已经训练好的网络模型中,这样不仅缩短了训练时间,而且得到了较高的精度。

## 1.2 迁移学习

迁移学习于 1990 年出现在机器学习领域<sup>[16-17]</sup>。迁移学习的实质就是运用已有的知识解决相关问题的一种方法,最终实现知识在相关领域之间的迁移。在深度学习的过程中,为了克服实验数据样本过少,导致模型泛化能力不足,网络出现过拟合的现象,需引入迁移学习。采用迁移学习需要注意 2 个问题:一是新的学习中需要识别的类别在预训练模型中训练过;二是预训练模型应该具有足够的泛化能力。在计算机视觉和遥感影像处理领域,迁移学习的应用是很常见的,迁移学习能够优化网络训练模型,减少数据集训练的时间,解决卷积神经网络在训练时样本较少的问题,迁移学习的广泛应用在一定程度上也能扩大卷积神经网络的应用领域。一般来说,按照迁移学习研究的内容不同,将迁移学习分为 4 种<sup>[18]</sup>:以实例为研究对象、以特征为研究对象、以参数为研究对象以及基于关联规则的迁移学习。迁移学习示意图如图 3 所示。

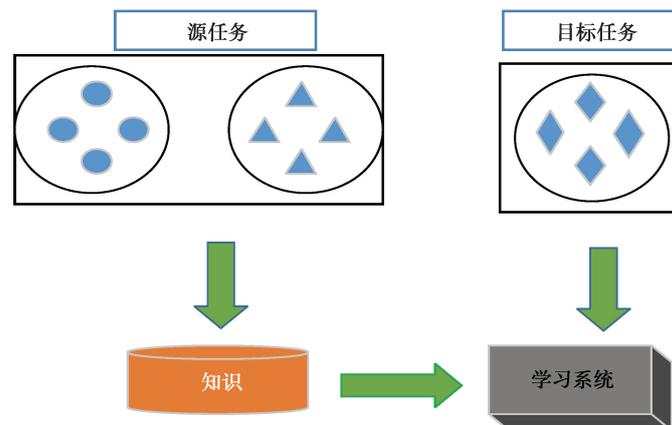


图 3 迁移学习示意图

Fig 3 Schematic diagram of migration learning

## 2 实验结果及分析

### 2.1 数据来源

数据来源于日本东北大学情报科学研究所,是日本某一地区海啸前后的全景街区影像。该数据获取使用与 GPS 数据匹配的车辆,其具体做法是在一辆车上安装全方位摄像头,通过车顶上的 GPS 传感器,在一个城市的街道上相隔一段时间,采集 2 次数据,得到 2 组全方位街区影像。分别在海啸前和海啸后进行数据采集,得到如图 4 所示的全景街区影像。选取 2 组原始影像作为实验数据,其中图 4 表示的是典型的全景街区影像,而图 5 表示的是空旷郊区的影像,选取这 2 组影像,目的是验证方法对不同类型的全景街区影像都适用。图 4、图 5 中 2 组影像的大小都为  $1\ 024\ \text{像素} \times 224\ \text{像素}$ ,通过目视判读,可以看出图 4 中包括了建筑物、天空、电线杆、道路、车辆等地物。图 5 中包括了建筑物、天空、车辆、空地等基本地物,从影像中可以看出,不同的拍摄环境及拍摄条件,得到的同一地区影像的光谱特征、纹理特征存在较大差异,这一现象给实验带来巨大挑战。实验选择最大似然法、SVM 作为 Segnet 的对比实验。



图 4 第一组实验数据

Fig. 4 Data of the first experiment

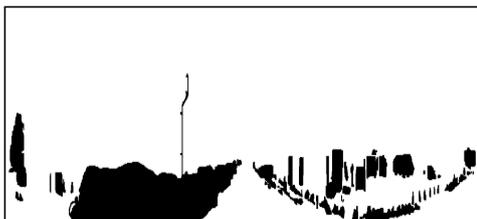


图 5 第二组实验数据

Fig 5 Data of the second experiment

### 2.2 实验结果分析

研究首先采用机器学习的方法对原始图像进行变化检测:基于 ENVI 对 2 组影像进行处理,分别采用了最大似然法和支撑向量机的方法对原始数据进行分类,将得到的分类数据相减,得到的结果如图 6(b)、6(c)和图 7(b)、7(c)所示;采用提出方法得到的变化检测结果如图 6(d)和图 7(d)所示。其中图 6 为原始数据一的变化检测结果图,图 7 为原始数据二的变化检测结果图。然后将相减结果与参考图 6(a)、图 7(a)分别进行精度评价。



(a) 变化参考图



(b) 最大似然法变化结果图

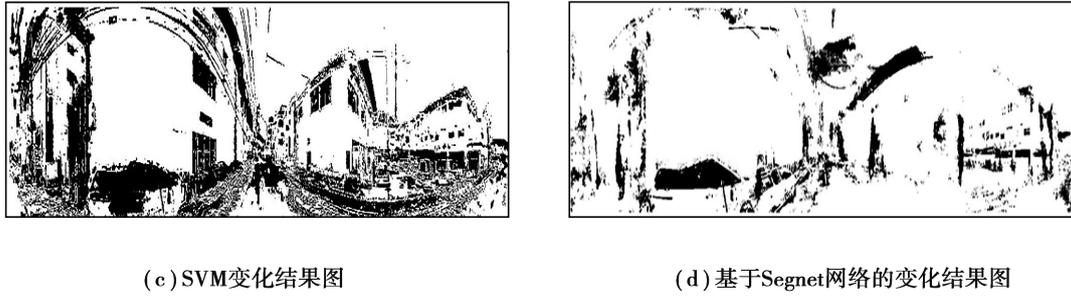


图 6 第一组实验结果图

Fig 6 The first set of experimental results

对于第一组实验结果,最大似然和 SVM 方法的精度分别为 65.1% 和 72.1%, 提出方法的精度为 81.4%。从变化的结果图中可以看出:与参考图对比,最大似然法和 SVM 的方法对建筑物墙体和道路的分类比较差,尤其是影像最左边的墙体及墙体上的窗户,都出现了错分现象,道路的分类主要存在的问题是边界线没有提取出来,仍然有错分的情况。对比图 4,可以看出造成这种现象最主要的原因是光谱差异,图 4 中变化前和变化后的影像光谱差异明显,并且主要体现在建筑物和道路上。造成这种现象的原因可能是数据获取时天气、光照等条件存在差异。而提出的方法对于建筑物和天空的分类结果比较好,一个比较大的问题是在变化后的影像中,对于车辆的提取结果较差,但是总体精度有所提高。

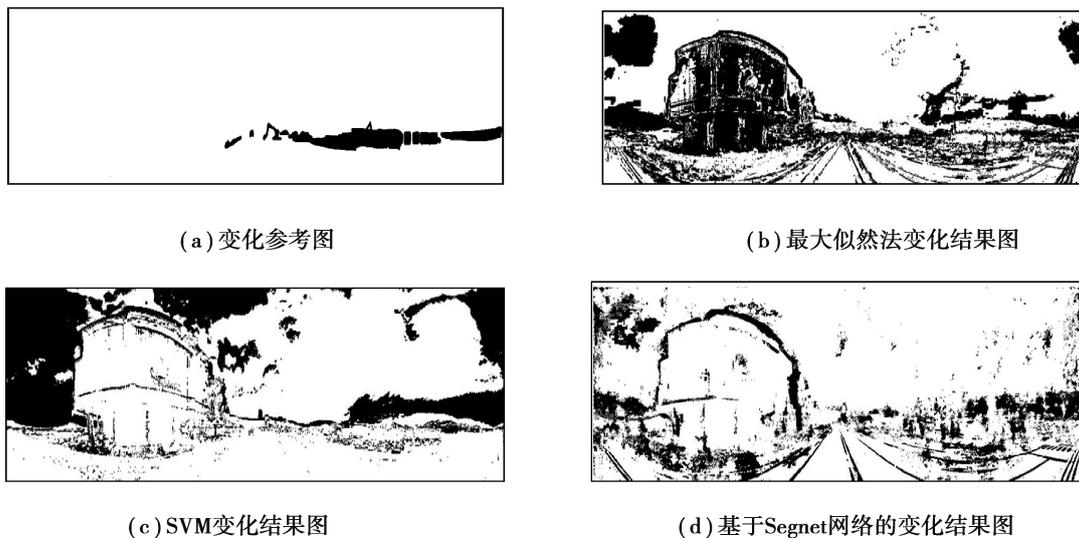


图 7 第二组实验结果图

Fig 7 The second set of experimental results

对于第二组实验结果,最大似然和 SVM 2 种方法的精度分别为 66.5% 和 70.6%, 提出的方法精度为 82.2%。从变化的结果图中看出:与参考图对比,最大似然法存在的问题是对建筑物的分类结果较差,基本整个建筑物都被错分为其他地物,主要原因是纹理信息比较相像。SVM 方法的优势在于对道路的提取有很好的效果,整个道路都被提取出来,且没有出现噪声和碎小的图斑,但是整个天空出现了较大部分错分的情况。参照图 5,可以初步推断造成这种现象的原因是变化前后的影像中,天空的光谱差异较明显。对于提出的方法,分类结果明显改善,但是也存在对地物边界识别不明显的问题,尽管如此,提出方法较最大似然法和 SVM 的变化检测精度有较大提高。根据以上变化检测结果,可以得出结论:虽然机器学习方法在遥感影像变化检测中应用也较广泛,但是仍然不适用于全景街区影像的变化检测。

研究采用漏检率、错检率以及总体精度作为精度评价的指标。其中漏检率指实际变化了,但是被检测为未变化的像元数占总像元的比例;错检率是指实际未变化,但被检测为变化的像元数站总像元数的比例;总

体精度指正确变化的像元数占总像元数的比例<sup>[19]</sup>。研究方法与机器学习的方法相比,其分类的精度和变化检测的精度都得到了明显提高。在第一组实验数据中,基于 Segnet 网络的变化检测方法较其他 2 种方法,精度分别提高了 16.3%和 9.3%;在第二组实验数据中,基于 Segnet 网络的变化检测方法较其他 2 种方法,精度分别提高了 15.7%和 11.6%。说明 Segnet 网络适用于全景街区影像的变化检测,Segnet 网络在全景街区影像变化检测中的应用,为街道违建、违章、违停排查,灾后评估提供了有力的理论支撑,对于城市的科学规划和辅助政府决策作出了重大贡献,其对比结果如表 1 和表 2 所示。

表 1 第一组实验结果精度对比

Table 1 Comparison of the accuracy of the first set of experimental results %

方法	漏检率	错检率	总体精度
最大似然法	4.9	30.0	65.1
SVM	5.7	22.2	72.1
Segnet	10.6	8.0	81.4

表 2 第二组实验结果精度对比

Table 2 Comparison of the accuracy of the second set of experimental results %

方法	漏检率	错检率	总体精度
最大似然法	1.5	32.0	66.5
SVM	2.6	26.7	70.6
Segnet	2.5	15.3	82.2

第一组实验结果中最大似然法、支持向量机、基于 Segnet 网络的漏检率分别为 4.9%、5.7%、10.6%,错检率分别为 30.0%、22.2%、8.0%;第二组实验结果中最大似然法、支持向量机、基于 Segnet 网络的漏检率分别为 1.5%、2.6%、2.5%,错检率分别为 32.0%、26.7%、15.3%。

### 3 结 论

针对采用传统方法难以得到高精度的全景街区变化信息的问题,提出了一种基于 Segnet 和迁移学习的全景街区影像变化检测方法。在实验过程中,实验数据主要是 2 组变化前后的全景街区影像,并采用了最大似然法以及 SVM 作为对比实验,得出以下结论:

1)从实验结果图中可以看出,道路和天空的变化信息的提取精度相对较高,而建筑物由于受纹理信息复杂性和相邻地物的干扰,其检测的精度相对较低。

2)2 组实验中最大似然法、SVM、Segnet 的总体精度分别为 65.1%、72.1%、81.4%和 66.5%、70.6%、82.2%,提出方法较最大似然法和支持向量机的方法精度明显提高。在灾后评估、街区违建违停排查、城市道路合理规划、土地利用变化等领域有着重要的作用。

3)基于 Segnet 神经网络和迁移学习的思想,通过对已有训练集的改进,使最终精度较机器学习的方法有了很大提高,但是并没有达到最佳效果。针对数据集重新制定训练集,并对数据重新进行实验,以得到更高精度是下一步将开展的工作。

#### 参考文献:

- [1] 吴辉. 全景影像的优化生成及其并行处理方法[D]. 南京: 南京师范大学, 2014.  
Wu H. The optimized generation and parallel processing methods of panoramic images[D]. Nanjing: Nanjing Normal University, 2014. (in Chinese)
- [2] 眭海刚, 冯文卿, 李文卓, 等. 多时相遥感影像变化检测方法综述[J]. 武汉大学学报·信息科学版, 2018, 43(12):

- 1885-1898.
- Sui H G, Feng W Q, Li W Z, et al. Review of change detection methods for multi-temporal remote sensing imagery[J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1885-1898. (in Chinese)
- [ 3 ] Wu C, Zhang L F, Zhang L P. A scene change detection framework for multi-temporal very high resolution remote sensing images[J]. Signal Processing, 2016, 124: 184-197.
- [ 4 ] 轩永仓. 基于全卷积神经网络的大田复杂场景图像的语义分割研究[D]. 杨凌: 西北农林科技大学, 2017.
- Xuan Y C. Research on the semantic segmentation of complex scene image of field based on fully convolutional networks [D]. Yangling: Northwest A & F University, 2017. (in Chinese)
- [ 5 ] Wu C, Zhang L P, Du B. Kernel slow feature analysis for scene change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(4): 2367-2384.
- [ 6 ] Amin A M, Liu Q, Wang Y. Convolutional neural network features based change detection in satellite images[C]//First International Workshop on Pattern Recognition 2016. Tokyo, Japan: [s.n.], 2016: 181-186.
- [ 7 ] 刘文涛, 李世华, 覃驭楚. 基于全卷积神经网络的建筑物屋顶自动提取[J]. 地球信息科学学报, 2018, 20(11): 1562-1570.
- Liu W T, Li S H, Qin Y C. Automatic building roof extraction with fully convolutional neural network[J]. Journal of Geo-Information Science, 2018, 20(11): 1562-1570. (in Chinese)
- [ 8 ] 魏杨, 毕秀丽, 肖斌. 基于区域卷积神经网络的农业害虫检测方法[J]. 计算机科学, 2018, 45(S2): 226-229, 233.
- Wei Y, Bi X L, Xiao B. Agricultural insect pest detection method based on regional convolutional neural network[J]. Computer Science, 2018, 45(S2): 226-229, 233. (in Chinese)
- [ 9 ] 邓国徽, 高飞, 罗志鹏. 基于改进的全卷积神经网络高分遥感数据语义分割研究[C]//第四届高分辨率对地观测学术年会论文集. 武汉: [s.n.], 2017: 13.
- Deng G H, Gao F, Luo Z P. Research on semantic segmentation of high-resolution remote sensing data based on improved full convolutional neural network[C]// proceedings of the 4th annual conference on high-resolution earth observation. Wuhan: [s.n.], 2017: 13. (in Chinese)
- [10] Audebert N, Saux B L, Lefèvre S. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks[C]//Asian conference on computer vision. Cham: Springer, 2016: 180-196.
- [11] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? [C]//Advances in neural information processing systems, NIPS 2014. Montreal, Quebec, Canada: MIT Press, 2014: 3320-3328.
- [12] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [13] 张庆辉, 万晨霞. 卷积神经网络综述[J]. 中原工学院学报, 2017, 28(3): 82-86, 90.
- Zhang Q H, Wan C X. Review of convolutional neural networks[J]. Journal of Zhongyuan University of Technology, 2017, 28(3): 82-86, 90. (in Chinese)
- [14] Boureau Y L, le Roux N, Bach F, et al. Ask the locals: Multi-way local pooling for image recognition[C]// 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2651-2658.
- [15] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [16] Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, IEEE: 2010: 2528-2535.
- [17] Robert C. Machine learning, a probabilistic perspective[J]. Chance, 2014, 27(2): 62-63.
- [18] 王惠. 迁移学习研究综述[J]. 电脑知识与技术, 2017, 13(32): 203-205.
- Wang H. Overview of migration learning research[J]. Computer Knowledge and Technology, 2017, 13(32): 203-205. (in Chinese)
- [19] 刘莹, 李强. 融合多特征的高分辨率遥感影像震害损毁建筑物检测[J]. 测绘与空间地理信息, 2018, 41(6): 61-64.
- Liu Y, Li Q. Damaged building detection from high resolution remote sensing images by integrating multiple features [J]. Geomatics & Spatial Information Technology, 2018, 41(6): 61-64. (in Chinese)