

DOI:10.19651/j.cnki.emt.2211910

采用样本自动选择的建筑物遥感场景分类方法^{*}

郭玉文 孙立双 谢志伟 史振国

(沈阳建筑大学交通与测绘工程学院 沈阳 110000)

摘要: 日前遥感图像建筑物场景分类方法多采用人工标注样本,标注过程需要大量时间。针对该问题,提出了一种采用样本自动选择的高分遥感建筑物场景分类方法。首先,建立光谱特征、几何特征和深度特征的多维高分遥感影像影像对象特征空间;其次,采用决策树初步提取建筑物,构建建筑物场景密度直方图;然后,采用自然间断法对建筑物密度分级,并采用比例法分别在每类场景中提取部分场景图像作为训练样本;最后,采用 ResNet50 网络对建筑物场景分类。以辽宁省沈阳市浑南区为研究区域和 Google Earth 遥感图像为实验数据,实验结果表明本文方法能够实现非监督场景分类,总体分类精度和 Kappa 系数分别为 0.89 和 0.82,较原有样本选择方法分类精度提高了 3% 和 8%。

关键词: 建筑物场景分类;样本自动选择;深度特征;高分遥感影像

中图分类号: P237 **文献标识码:** A **国家标准学科分类代码:** 420.42020

Automatic sample selection method for building remote sensing scene classification

Guo Yuwen Sun Lishuang Xie Zhiwei Shi Zhenguo

(School of Transportation and Geomatics Engineering, Shenyang Jianzhu University, Shenyang 110000, China)

Abstract: At present, most of the building scene classification methods in remote sensing images use manual annotation method, which requires a lot of time. To solve this problem, this paper proposes a high resolution remote sensing building scene classification method using automatic sample selection. Firstly, the feature space of multi-dimensional high-resolution remote sensing image with spectral features, geometric features and depth features is established. Secondly, the buildings were initially extracted by decision tree, and the scene density histogram of buildings was constructed. Then, the natural discontinuity method was used to classify the building density, and the proportion method was used to extract some scene images from each type of scene as training samples. Finally, ResNet50 network is used to classify building scenes. Taking Hunnan District of Shenyang City, Liaoning Province as the study area and Google Earth remote sensing images as the experimental data, the experimental results show that the proposed method can achieve unsupervised scene classification, and the overall classification accuracy and Kappa coefficient are 0.89 and 0.82, respectively, which are improved by 3% and 8% compared with the original sample selection method.

Keywords: building scene classification; automatic sample selection; depth feature; high resolution remote sensing image

0 引言

通过高分遥感图像可以提取大量的场景信息^[1],而建筑物场景可以提供包含例如城市建成现状、城市地表空间结构的重要信息。因此,面向建筑物的高分遥感图像场景分类在现代城市结构建设、中心城区用地规划以及城市区域生态、气候环境中具有重要指导意义^[2]。

为了快速、准确识别图像中的场景语义信息,目前国内

外研究学者主要采用基于中、低层特征和基于高层特征的两种不同方法。在分别采用低、中、高层等特征进行分类时,有学者提出基于局部颜色直方图识别建筑物的分层方法以及基于空间纹理和颜色特征的稀疏表示的建筑物识别方法^[3-4];还有学者将建筑物表示为特征空间中的一组“稀疏”特征向量,利用稀疏标记代表建筑物提取的信息^[5];在使用特征组合方法进行分类时,有学者提出分别对浅层-中层-深层-多层的特征进行特征选择与组合,也有学者基

收稿日期:2022-11-01

^{*} 基金项目:国家自然科学基金(42101353)、教育部人文社会科学研究一般项目(21YJC790129)、辽宁省教育厅基本科研项目(LJKMZ20220946)资助

于低维卷积神经网络特征提取的多核支持向量机场景分类算法,提取低维卷积神经网络感知器中的不同特征进行加权融合,还有学者提出基于多结构深度特征融合的分类方法,分别使用3种主流的卷积神经网络(convolutional neural networks, CNN)作为特征提取器,从图像中提取深度特征,采用深度特征融合网络对这些特征进行融合,实现遥感影像场景分类^[6-8]。基于图像的低、中、高层特征可以检测出建筑物,但在使用以上方法时,大多采用人工标注样本的方法,标注过程需要大量时间。

在使用联合特征进行样本标注方面,有学者提出深度联合光谱-空间特征表示法和基于 AlexNet 的特征学习网络,提升特征的多层次表示能力并降低样本量要求^[9-10];也有学者提出基于深度学习的光谱空间联合特征的高光谱图像分类方法,结合三维 CNN 和灰度共生矩阵^[11](grey-level co-occurrence matrix, GLCM)算法,通过充分利用每个像素相邻未标记样本所包含的光谱空间信息,深度挖掘光谱空间信息,提高分类准确性^[12];还有学者提出基于特征相似度,训练线性分类器,根据分类置信度,提取高置信度样本分配伪标签,构造标注样本集^[13];有学者将标注问题转化为多标签学习问题,提出基于深度卷积神经网络和多标签数据增强的端到端自动图像标注方法,使用深度 CNN 结构进行自适应特征学习^[14]。但以上方法只是减少对样本的标记要求,在选择 CNN 训练样本进行场景分类方面均没有达到自动化或半自动化水平。

针对以上问题,本研究提出采用联合低层特征和深度特征进行样本自动选择的高分遥感建筑物场景分类方法,从而在不降低建筑物提取模型精度的情况下,显著减少人工和时间成本的投入。

1 采用样本自动选择的高分遥感建筑物场景分类方法

针对以上问题,本研究提出采用联合低层特征和深度特征进行样本自动选择的高分遥感建筑物场景分类方法,从而在不降低建筑物提取模型精度的情况下,显著减少人工和时间成本的投入。本研究整体技术流程如图1所示。

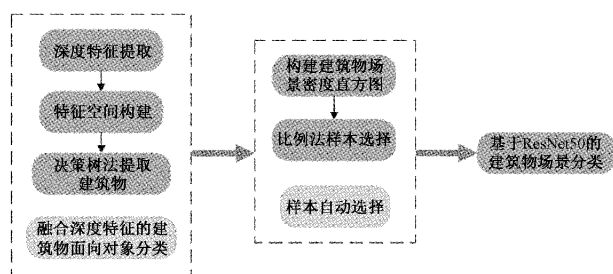


图1 技术流程

1.1 融合深度特征的建筑物面向对象分类

1) 深度特征提取

由于 ResNet^[15] 网络结构后面特征层的内容会有一部

分由前面某一层的线性贡献,层数越深,性能越佳,且网络模型层数固定,所以本研究选用 ResNet50 网络模型作为图像特征提取器,来提取图像深度特征。遥感图像的 R、G、B 三通道各包含不同的特征信息,将 RGB 图像作为输入层,通过 ResNet50 网络模型五个阶段的卷积、池化以及后续非线性激活后学习到的特征作为输出层,得到遥感图像三通道的深度特征。

在 CNN 中,修正线性单元(rectified linear unit, ReLU)作为神经元的激活函数,能够去除卷积结果中的负值,保留正值,实现神经元稀疏。通过 ReLU 稀疏后的模型能够更好地挖掘建筑物的深度特征,拟合训练样本。因此,本研究选用 ReLU 层作为提取图像深度特征的关键层,使得周围地物特征减弱,建筑物特征得以强化。

2) 特征空间构建

根据地物特点,建立基于几何特征、纹理特征的低中层特征和基于深度特征的高层特征的特征空间。其中,几何特征包括紧密度(compactness)、长宽比(length-width)、椭圆拟合(elliptic fit)以及形状指数(shape index);纹理特征包括基于 GLCM 的均值(mean)、角二阶矩(angular second moment, ASM)、熵(entropy, ENT)、对比度(contrast)、反差分矩矩阵(inverse differential moment, IDM)、相关性(correlation)、同质性/逆差矩(homogeneity)、非相似性(dissimilarity);深度特征则为基于 ResNet50 网络提取的图像特征,将提取的 ReLU 层的深度特征作为波段信息加入图像,形成完整特征空间。

3) 决策树法提取建筑物

将研究区内遥感影像进行预处理后进行多尺度分割,对分割好的地物,利用决策树根据构建的特征空间选用最优特征组合进行分类。使用 Sentinel-2 数据,利用归一化植被指数(normalized difference vegetation index, NDVI)、归一化水指数(normalized difference water index, NDWI)对研究区内植被和水体进行提取后,融合无人机影像数据,利用构建的特征空间,通过决策树对建筑物进行精确提取。本研究将地物分为建筑物(Building)、农田(Farmland)、林地(Forest)、道路(Road)、水体(Water),地物类型以《土地利用现状分类》GB/21010-2017 为标准,此分类体系为现行标准,可以满足研究需求。

1.2 样本自动选择

1) 构建建筑物场景密度直方图

在场景分类时,采用分辨率为 1 m 的高分辨率遥感图像,根据棋盘分割方法划分 200 pixels×200 pixels 的遥感图像为单位场景。为了使建筑物场景数据分布可视化,对 CNN 训练样本选择更合理,本研究采用密度图的形式来实现。通过统计各场景内提取到的建筑物总面积占划分单位场景面积的比例情况,得到各建筑物遥感场景的密度,对其进行统计分析,得到建筑物场景密度占比直方图。为了使密度分布情况显示更直观,设置直方图组距为 0.05,即建

建筑物场景密度的跨度为 0.05,对研究区内各级密度场景数量更好掌握。

2) 比例法样本选择

利用建筑物密度占比直方图的分布信息,综合场景图像数量以及各比值变化率,采用自然断点法划定研究区域内建筑物遥感场景图像的分类级别密度。在实验区总体要进行分类的 3 428 张建筑物遥感场景中,以 1:5 的比例选择出各类型密度的 686 张建筑物场景作为训练样本。由于研究区整体场景数量在万张以内,数据集较小,所以采用简单交叉验证的方式,只设置训练集与测试集,不设置验证集。同时,为了降低类间误分率,使模型分类效果获得最佳表现,各类型密度建筑物场景训练集应尽量均衡,所以训练样本比例设置为 1:1:1。

2 基于 ResNet50 的建筑物场景分类

ResNet 采用残差模块 ShortCut connection 跳跃式的结构,打破了传统 CNN 只能逐层传递参数的方法,可以越过中间的某几层直接将参数传递给后面的层,改善了网络退化的问题,而且具有恒等映射层,使得网络加深时不会带来训练集误差的上升,因此本研究选用 ResNet50 网络模型来进行建筑物场景分类。ResNet50 网络模型的残差模块结构如图 2 所示。

将 ResNet50 网络在 ImageNet 上学习到的特征提取能力迁移到遥感图像场景分类任务中,去除其全连接层,将

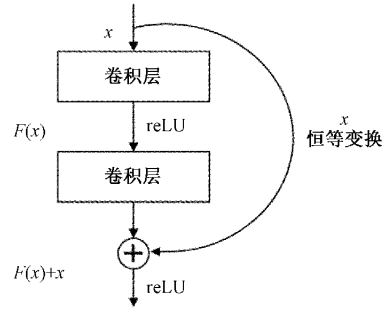


图 2 残差模块结构

剩下的网络当作一个固定的特征提取机,通过在全连接层之前加入 Dropout 防止过拟合,另外搭建适合本研究的 ResNet50 网络模型的全连接层,对 ResNet50 网络模型进行微调。根据场景内建筑物大体结构被提取出后划定的样本,微调后的 ResNet50 网络模型即可对场景的密度级别有较好判断,在达到 80% 的建筑物被提取出后即可满足实验精度要求。最后,引入迁移学习的思想,将按比例自动选择的 CNN 训练样本,使用微调后的 ResNet50 网络模型对研究区内建筑场景进行训练并分类。残差网络的跳跃结构使得每次工作的网络并不深,某种意义上,可以看成是一些不同的浅层神经网络的组合。对建筑物密度级别有较好的预判能力,解决了建筑物遥感场景的分类问题。基于迁移学习的遥感图像场景分类流程如图 3 所示。将研究区内场景图像放入微调的 ResNet50 网络中,得到场景分类实验结果。

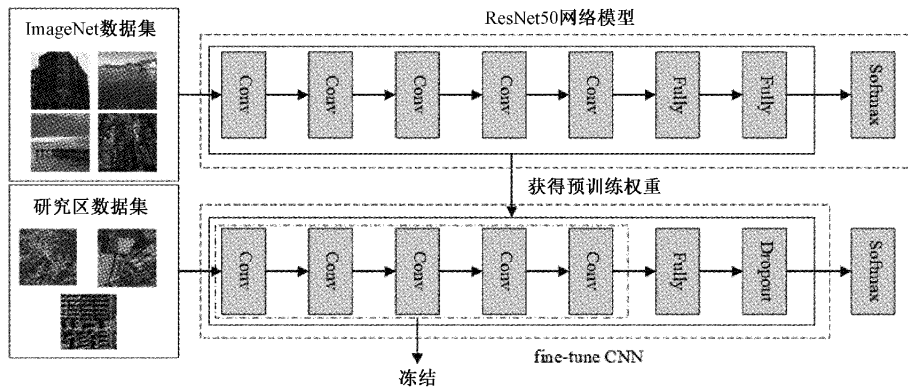


图 3 基于迁移学习的遥感图像场景分类流程结构

3 实验与分析

3.1 深度特征提取

利用 ResNet50 网络模型提取 ReLU 各层深度特征,在深度特征图中,建筑物轮廓与周围地物具有显著区别,强化了建筑物的特征。提取的建筑物深度特征如图 4 所示。

3.2 决策树建筑物提取

在对实验数据进行多尺度分割后,通过决策树对数据利用传统图像特征等构建的规则集进行解译,经传统低层

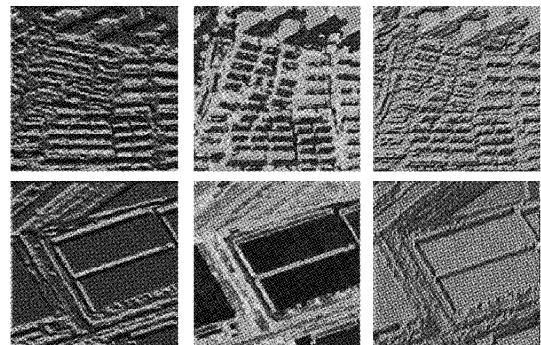


图 4 建筑物深度特征

图像特征等筛选后的建筑物提取细节如图 5 所示。在将提取的深度特征加入特征空间后,联合低层-深度特征对建筑物提取的细节如图 6 所示。

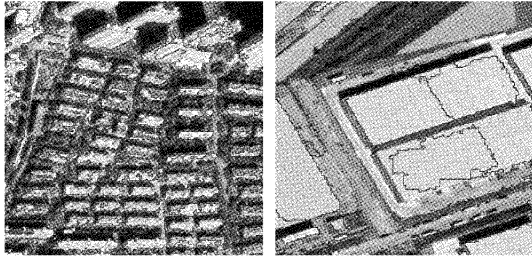


图 5 低层特征提取的建筑物

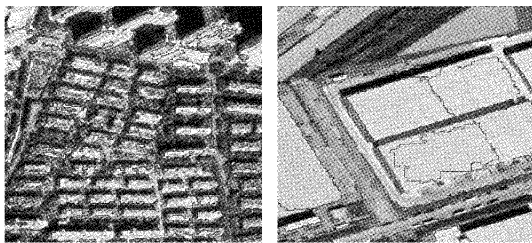


图 6 联合低层-深度特征提取的建筑物

和未添加深度特征提取的建筑物对比,加入深度特征后提取的建筑物更为细致,可以筛除混淆的农田、部分道路地块,建筑物提取效果更好,场景内建筑物的密度计算更为精确。

混淆矩阵,是机器学习中进行精度评价的一种标准格式,以 n 行 n 列的矩阵形式将数据集中的真实的类别与分类模型预测的类别进行汇总。其中, n 为类别的数量,矩阵的行表示预测值,矩阵的列表示真实值,矩阵的主对角元素为正确分类数,对角线以外的元素为分类相对于真实参考数据的错误分类数。

总体分类精度(overall accuracy, OA)为被正确分类的数据个数与总体数据个数的比值,OA 可以直接反映出分类正确的比例。

Kappa 系数是用于一致性检验的指标,用于衡量模型预测结果和实际分类结果是否一致的分类效果,Kappa 系数在 $-1 \sim 1$,计算结果通常落在 $0 \sim 1$,可分为 5 组来表示不同级别的一致性: $0.0 \sim 0.20$ 极低一致性、 $0.21 \sim 0.40$ 一般一致性、 $0.41 \sim 0.60$ 中等一致性、 $0.61 \sim 0.80$ 高度的一致性、 $0.81 \sim 1.0$ 几乎完全一致。

在决策树提取建筑物后,在研究区内随机均匀设置 200 个验证点,与人工判定的场景分类结果进行对比,利用混淆矩阵对建筑物提取结果进行精度评价,验证点混淆矩阵如表 1 所示。

根据混淆矩阵,计算出 OA 为 0.91, Kappa 系数为 0.81,说明决策树提取建筑物的精度可以满足本研究要求。

表 1 验证点混淆矩阵

验证点	预测		总计
	建筑物	非建筑物	
建筑物	170	10	180
实际非建筑物	8	12	20
总计	178	22	200

3.3 样本自动选择

根据各场景内建筑物密度占比直方图,利用自然断点法将密度为 $0.04 \sim 0.13$ 范围内的图像划分为低密度场景, $0.14 \sim 0.20$ 范围内的图像划分为中密度场景, $0.21 \sim 0.64$ 范围内的图像划分为高密度场景。以 $1:5$ 的比例从 3 428 张建筑物遥感场景中,选择出各类型密度的 686 张建筑物场景作为训练集。其中,低密度建筑场景:中密度建筑场景:高密度建筑场景按 $1:1:1$ 的分布,从而实现每类建筑物的样本自动选择。场景内建筑物占比直方图如图 7 所示。

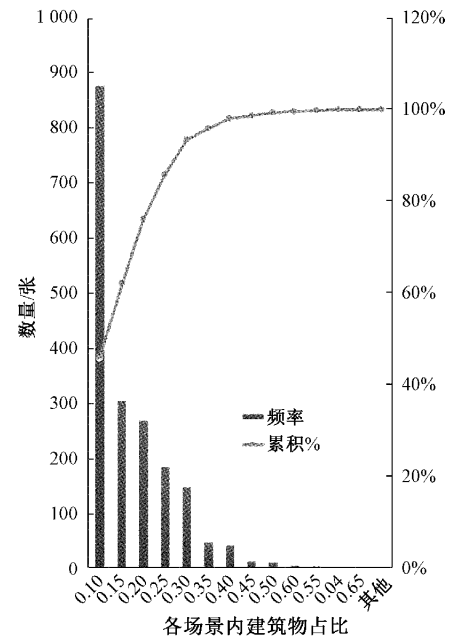


图 7 各场景内建筑物占比直方图

在研究区内创建 217 个随机均匀分布的验证点,对三类密度场景图像进行精度验证,通过计算混淆矩阵、总体精度、Kappa 系数等,进行精度评定。研究区内验证点混淆矩阵如表 2 所示。

表 2 验证点混淆矩阵

验证点	高密度	中密度	低密度	总计
	建筑	建筑	建筑	
高密度建筑	30	0	3	33
中密度建筑	2	4	5	11
低密度建筑	1	0	172	173
总计	33	4	180	217

根据混淆矩阵,计算出 OA 为 0.95, Kappa 系数为 0.84,说明自动选择样本的精度可以满足本研究要求。

为进一步证明各验证点可靠性,特从以下几方面来进行进一步的精度评定:虚分误差,指被分为用户感兴趣的类,而实际上属于另一类的数据,虚分误差显示在混淆矩阵的行里面;漏分误差,指本属于地表真实分类,但没有被分到相应类别中的数据,漏分误差显示在混淆矩阵的列里面;制图精度,指正确分为某一类的数据与某一类真实参考总数的比例;用户精度,指正确分到某一类类的数据与整个数据集中 A 类的数据总数比例。各验证点场景密度分类精度评定表如表 3 所示。

表 3 各验证点场景密度分类精度评定表

密度级别	虚分误差	漏分误差	制图精度	用户精度
高密度建筑	0.09	0.09	1.00	0.36
中密度建筑	0.64	0	0.91	0.91
低密度建筑	0.01	0.04	0.96	0.99

3.4 场景分类结果及精度评价

方法 1) 使用人工选择训练样本,放到微调的 ResNet50 网络内进行训练,并进行测试,得到浑南区建筑物遥感图像场景分类的训练集的准确度、训练集的损失值、验证集的准确度、验证集的损失值如图 8 所示,使用人工选择的 CNN 训练样本场景分类实验结果如图 9 所示。

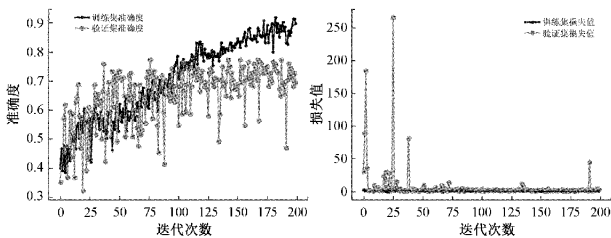


图 8 使用人工选择的 CNN 样本训练的准确度和损失值

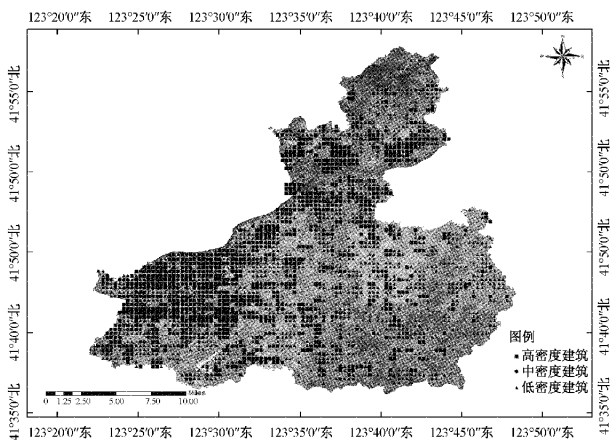


图 9 使用人工选择的 CNN 样本场景分类实验结果图

其中,训练集的准确率为 0.90,损失值为 0.31,验证集

的准确率为 0.70,损失值为 0.72。

方法 2) 使用低层特征选择训练样本,放到微调的 ResNet50 网络内进行训练,并进行测试,得到浑南区建筑物遥感图像场景分类的训练集的准确度、训练集的损失值、验证集的准确度、验证集的损失值如图 10 所示。

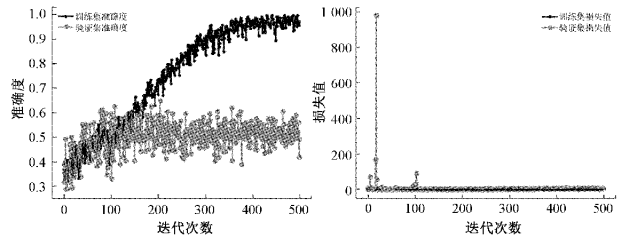


图 10 使用低层特征选择的 CNN 样本训练的准确度和损失值

其中,训练集的准确率为 0.97,损失值为 0.08,验证集的准确率为 0.42,损失值为 0.82。使用低层特征选择的 CNN 训练样本场景分类实验结果如图 11 所示。

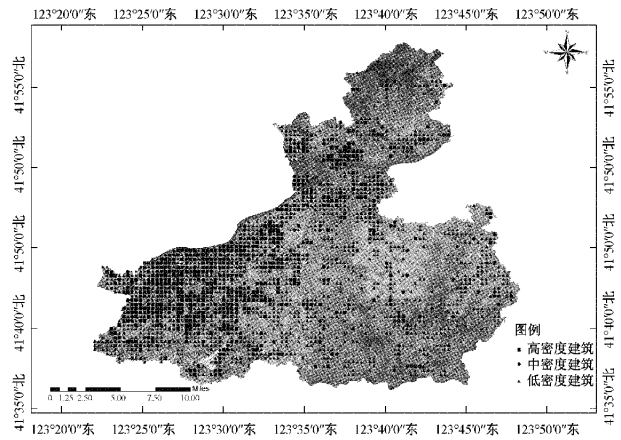


图 11 使用低层特征选择的 CNN 样本场景分类实验结果图

方法 3) 使用联合低层-深度特征自动提取训练样本,放到微调的 ResNet50 网络内进行训练,并进行测试,得到浑南区建筑物遥感图像场景分类的训练集的准确度、训练集的损失值、验证集的准确度、验证集的损失值如图 12 所示。

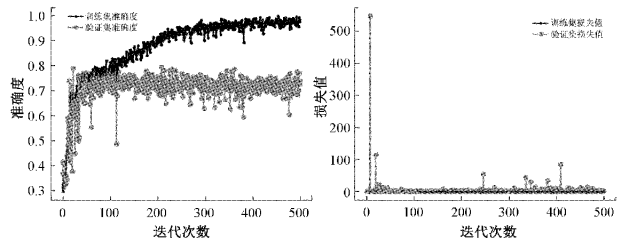


图 12 联合低层-深度特征选择的 CNN 样本训练的准确度和损失值

其中,训练集的准确率为 0.97,损失值为 0.09,验证集的准确率为 0.65,损失值为 0.04。

在加入深度特征后,使用联合低层-深度特征的验证集准确度有所提高,同时验证集的损失值下降,说明模型训练效果较好,可以用于场景分类。添加深度特征选择的 CNN 训练样本场景分类实验结果如图 13 所示。

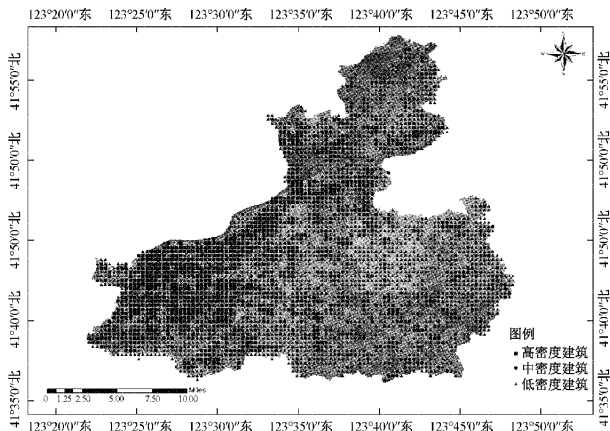


图 13 联合低层-深度特征选择的 CNN 样本场景分类实验结果图

将深度特征添加到构建的特征空间后,对研究区域的建筑物遥感场景识别更全面,说明用提前提取好的深度特征图作为选样本的重要因素,对建筑物密度进行判别是可行的。

使用 3 种不同方法选择样本后对浑南区建筑物进行场景识别,得到的混淆矩阵如表 4 所示。

表 4 使用不同方法选择样本的场景分类结果混淆矩阵

方法	预测	高密度建筑	中密度建筑	低密度建筑	总计
	实际				
1	高密度建筑	730	17	81	828
	中密度建筑	12	193	20	225
	低密度建筑	22	11	696	729
	总计	764	221	797	1 782
2	高密度建筑	604	71	15	690
	中密度建筑	30	28	66	124
	低密度建筑	24	53	944	1 021
	总计	658	152	1 025	1 835
3	高密度建筑	1 003	79	26	1 108
	中密度建筑	22	419	239	680
	低密度建筑	1	3	1 636	1 640
	总计	1 026	501	1 901	3 428

根据混淆矩阵,对使用 3 种不同方法选择样本的场景识别结果进行精度评定,精度评定表如表 5 所示。

从以上不同方法的实验结果及精度评定来看,使用传统人工选择样本的方法进行场景分类,其总体精度,Kappa 系数都是最好的,使用联合低层-深度特征选择样本的方法进行场景分类结果次之且与人工选择样本的方法进行场

表 5 使用不同方法选择样本的场景分类结果精度评定表

方法	评价指标			
	总体精度	Kappa 系数	虚检率/%	漏检率/%
1	0.91	0.85	31	30
2	0.86	0.74	97	98
3	0.89	0.82	48	33

景分类结果差距较小,而使用低层特征选择样本的方法进行场景分类结果最次且与人工选择样本的方法进行场景分类结果差距较大。同时,漏检率和错检率的情况与总体精度的优劣成正比。由此来看,本研究提出的使用自动选择 CNN 训练样本的方法进行场景分类在节省人工成本的同时能保证分类结果的精度,并且比仅使用低层特征选择样本的方法进行场景分类精度更高,效果更好。

4 结 论

本研究通过自动提取 CNN 训练样本对浑南区建筑物实现了空间分布密度识别,与使用低层特征选择 CNN 训练样本的方法进行遥感图像场景分类的结果对比,使用该样本选择方法,在分类精度方面,总体分类精度提高 3%,达到 0.89,Kappa 系数提高 8%,达到 0.82;同时与使用传统人工选择 CNN 训练样本的方法进行遥感图像场景分类的结果对比,两种样本选择方法在分类精度方面,总体精度与 Kappa 系数的水平接近,但是使用构建的联合低层-深度特征识别的建筑物遥感图像场景更为全面,且可省去人工标注样本的时间,说明所提出的训练样本自动选择方法可行有效。

参考文献

- [1] 李二珠. 遥感图像场景深度学习与应用研究[D]. 南京:南京大学, 2017.
- [2] 伍丝琪. 基于目标空间关系和多源地理数据的高分辨率遥感城市场景语义分类方法研究[D]. 武汉:武汉大学, 2019.
- [3] ZHANG W, KOSECKA J. Localization based on building recognition[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR05)-Workshops, 2005: 21-21.
- [4] LI B, SUN F, ZHANG Y J. Building recognition based on sparse representation of spatial texture and color features [J]. IEEE Access, 2019, 7: 37220-37227.
- [5] CHEN K, ZOU Z, SHI Z. Building extraction from remote sensing images with sparse token transformers[J]. Remote Sensing, 2021, 13(21): 4441.
- [6] 胡静文. 面向高分辨率遥感影像场景分类的层次化特征表达研究[D]. 武汉:武汉大学, 2019.
- [7] XUE W, DAI X, LIU L. Remote sensing scene

- classification based on multi-structure deep features fusion[J]. *IEEE Access*, 2020, 8: 28746-28755.
- [8] 耿万轩, 周维勋, 金双根. 基于 LDCNN 特征提取的多核 SVM 高分辨率遥感影像场景分类[J]. *测绘通报*, 2021, (8): 14-21.
- [9] LI D, YANG R, LI X, et al. Radar signal modulation recognition based on deep joint learning [J]. *IEEE Access*, 2020, 8: 48515-48528.
- [10] DENG C, XUE Y, LIU X, et al. Active transfer learning network: A unified deep joint spectral-spatial feature learning model for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(3): 1741-1754.
- [11] HARALICK RM, SHANMUGAM K, DINSTEN I H, et al. Textural features for image classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, (6): 610-621.
- [12] HAN Y, GAO Y, ZHANG Y, et al. Hyperspectral sea ice image classification based on the spectral-spatial-joint feature with deep learning [J]. *Remote Sensing*, 2019, 11(18): 2170.
- [13] 白坤, 慕晓冬, 陈雪冰, 等. 融合半监督学习的无监督遥感影像场景分类[J]. *测绘学报*, 2022, 51(5): 691-702.
- [14] KE X, ZOU J, NIU YJ. End-to-end automatic image annotation based on deep CNN and multi-label data augmentation[J]. *IEEE Transactions on Multimedia*, 2019, 21(8): 2093-2106.
- [15] 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*, 2016: 770-778.

作者简介

谢志伟, 博士, 副教授, 主要研究方向为遥感图像处理、城市遥感等。

E-mail: zwxrs@sjzu.edu.cn

郭玉文, 硕士, 主要研究方向为遥感图像处理。

E-mail: g381374963@163.com