

引用格式:Chen Min, Pan Jiawei, Li Jiangjie, *et al.* High Resolution Remote Sensing Image Construction Land Detection Combined with VGGNet and Mask R-CNN[J]. Remote Sensing Technology and Application, 2021, 36(2): 256-264.[陈敏, 潘佳威, 李江杰, 等. 结合VGGNet与Mask R-CNN的高分辨率遥感影像建设用地检测[J]. 遥感技术与应用, 2021, 36(2): 256-264.]
doi:10.11873/j.issn.1004-0323.2021.2.0256

结合VGGNet与Mask R-CNN的高分辨率 遥感影像建设用地检测

陈敏^{1,2}, 潘佳威², 李江杰², 徐璐², 刘加敏^{1,2}, 韩健³, 陈奕云^{2,4}

(1. 广州市城市规划勘测设计研究院, 广东 广州 510060;

2. 武汉大学资源与环境科学学院, 湖北 武汉 430079;

3. 广西华遥空间信息科技有限公司, 广西 南宁 530031;

4. 土壤与农业可持续发展国家重点实验室, 江苏 南京 210008)

摘要:针对当前多数深度学习模型只能对高分辨率遥感影像裁剪图片进行土地利用类型判别的问题,结合VGGNet与Mask R-CNN开展了智能化建设用地目标检测研究。在建立研究区4类土地利用类型遥感影像数据集的基础上,对比了VGGNet、ResNet和DenseNet 3种卷积神经网络模型的分类精度,选取分类效果最优的神经网络模型VGGNet与Mask R-CNN实现建设用地目标检测智能化。结果表明:①VGGNet、ResNet和DenseNet 3种卷积神经网络模型的分类精度分别为:97.44%、93.75%和95.13%,且VGG16模型迭代次数最少,训练时间相对较少;②Mask R-CNN阈值设置对目标检测精度有重要的影响,当阈值设定为0.3时,VGG16结合Mask R-CNN的联合模型对建设用地检测的标定框精度最高。同时联合模型比单一使用Mask R-CNN模型对建设用地检测有更高的准确率,并且表现出了更强的适应性和鲁棒性。

关键词:卷积神经网络(CNN);目标检测;影像分类;高分辨率遥感影像;建设用地

中图分类号:TP79 **文献标志码:**A **文章编号:**1004-0323(2021)02-0256-09

1 引言

随着遥感技术的快速发展和大量高分辨率遥感影像数据的涌现,遥感影像中所包含的信息日益丰富,已逐渐用于环境监测、资源管理、城市规划、土地利用等诸多领域精细化信息的获取。然而现有研究表明^[1-4],高分辨率遥感影像分类和目标检测面临着诸多挑战:一方面,高分辨率遥感影像对地面目标的细微差别能够做出更清晰地展示,使目标地物的类间差异有了一定减小,但与此同时类内差

异增大,从而加大了分类难度;另一方面,高分辨率遥感影像的波段相对较少,导致光谱信息欠缺,应用传统识别方法能够提取到的地物特征有限,这需要研究更加精确高效的遥感影像分类技术。

基于深度学习的高分辨率遥感影像分类和目标检测方法,是一种更加高效且准确的高分辨率遥感影像分类方法,它通过模拟人脑对遥感影像进行辨别,能够有效避免传统目视解译因人为主观因素带来的缺陷,并能够充分考虑到高分辨率遥感影像特有的复杂特性,对遥感影像进行快速、高效地分

收稿日期:2019-10-14;修订日期:2020-03-12

基金项目:国家重点研发计划“绿色宜居村镇技术创新”重点专项项目子课题“村镇发展潜力因子识别与指标信息快速获取技术”(2018YFD1100801-01)。

作者简介:陈敏(1994—),男,湖北咸宁人,硕士研究生,主要从事土地变化科学与可持续发展研究。E-mail:15927214729@163.com

通讯作者:韩健(1984—),男,广西柳州人,高级工程师,主要从事摄影测量、地理信息系统工程研究。E-mail:hanjianspace@126.com

类和目标检测,并对探索进行高分辨影像分类检测的新方法与新技术具有重要的理论意义^[5-6]。相比传统的影像分类方法,基于深度学习的各种模型在高分辨率遥感影像分类方面有更大的应用前景,且有效解决了传统方法通常只针对单一土地利用类型分类识别的问题^[7],其中 LeNet5、AlexNet、VGG-Net、GoogleNet、ResNet、DenseNet 等模型在对高分辨率遥感影像分类的过程中使用都较为广泛,相比传统方法普遍有更高的精度。针对高空间分辨率的遥感影像,深度学习模型对高分辨率影像分类及建设用地目标检测方面的适用性尚未被很好地评估。因此,本文旨在探索适用于高分辨率遥感影像的分类及目标检测模型,以期为将来的高分辨遥感影像分类提供借鉴与参考。

2 数据来源与研究方法

2.1 研究区

选择广西壮族自治区百色市的右江区作为研究区,右江区地处云、黔、桂三省(区)结合部,位于广西西部的珠江流域右江上游,是连接三省的交通枢纽和重要的边境物资集散地,作为大西南海上运输的主要通道,毗邻东盟市场,依靠着其得天独厚的区位、交通和政治优势,是广西的一大政治、经济、文化中心。根据第六次人口普查数据:右江区总人口 32.01 万人,辖区内分为 4 个镇、2 个街道、2 个乡和 1 个民族乡,土地面积达 3 713 km²,平地总面积 113.84 km²;水面面积 81.06 km²。右江区有林地

面积总共约为 28.4 万 hm²;其中耕地占 9.09%,面积约为 3.36 万 hm²;共有 3.04 万 hm²的未利用土地,约占总土地利用面积的 8.22%。右江区境内河网密布,水资源丰富,有 32 条河流集雨面积大于 25 km²,其中有 21 条超过了 50 km²。右江区区位图如图 1 所示。

2.2 数据与方法

2.2.1 数据来源

利用右江区第二次土地利用调查数据库中的土地利用矢量数据,并参考谷歌卫星影像地图上下载的高分辨率遥感影像瓦片数据,经过目视判读、修正、完善,形成研究区土地覆被数据。

由于数据集构建所需图斑的识别和采集量大,本次选择矢量图斑覆盖影像瓦片的分类提取数据方法,将土地利用分类矢量图斑覆盖到遥感影像瓦片上,用瓦片地图服务(TMS)的换算方法获取矢量图斑内的所有影像瓦片图片^[8]。此次研究只针对了研究区内的遥感影像数据进行采集分析,数据来源较为单一。为了保证模型能有较高的训练精度和较好的泛化能力,通过制定统一的筛选标准:影像完整清晰、目标地物类型特征明确且在图像中占比达到 80% 以上、地类表征正确,对获取的遥感影像数据进行人工筛选。最终结合制定的筛选标准,根据土地利用现状分类将遥感影像分为建设用地、耕地、林地和水域 4 类来选取试验数据。4 类实验数据实例如图 2 所示。

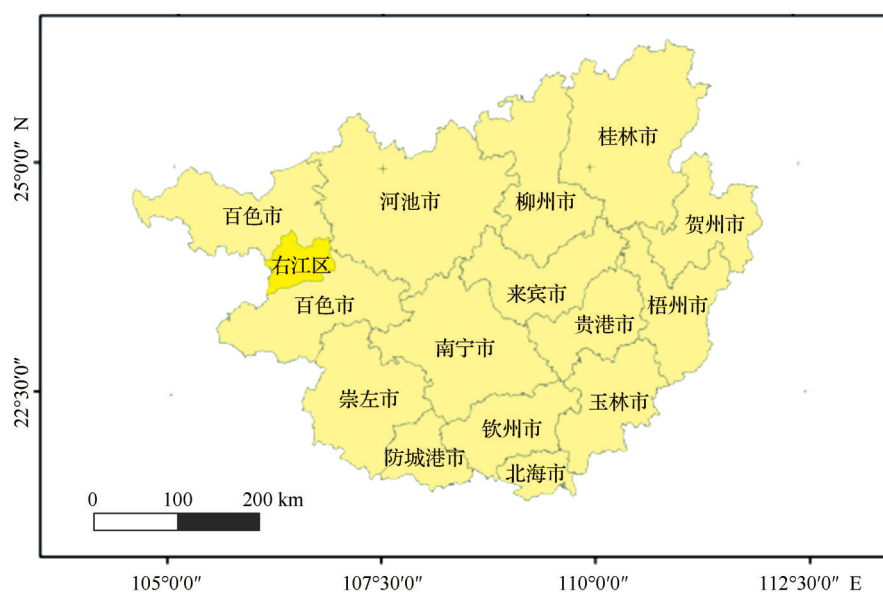


图1 右江州区位图

Fig.1 The location map of Youjiang district

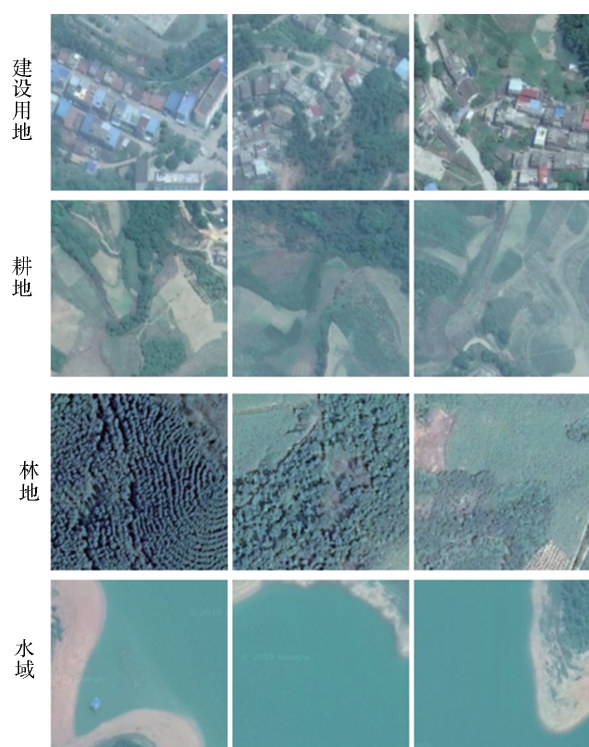


图2 影像数据分类实例

Fig.2 Image data classification example

2.2.2 数据标准化

深度学习神经网络模型通常要求所输入的待分类图像为同样分辨率的标准矩形,为了方便数据输入和模型训练,在进行网络训练之前,对原始影像数据进行规范化预处理,将其转换为 $224 \times 224 \times 3$ 的统一大小矩形图像数据用于下一步的模型训练。为了提升模型精度和收敛速度,提高网络的学习效率,使训练后的模型有更好的泛化效果,本文对原始图像进行了标准化处理。图像标准化的实质就是通过去均值的方式实现数据的中心化处理,标准化公式为:

$$\text{image_standardization} = \frac{x - \mu}{\max(\sigma, \frac{1.0}{\sqrt{N}})}$$

其中: μ 是图像的均值, x 表示图像的矩阵, σ 表示标准差, N 表示图像 x 的像素数量。

2.2.3 数据增强

由于成功的神经网络通常都包含大量的参数,为了能够获取最佳的参数组合,需要大量的数据进行训练。然而,右江区耕地面积较少且分布不均,难以获取足够的样本,因此需要对获取到的符合要求的样本集进行数据增强。一方面通过增加训练的数据量来提高模型的泛化能力;另一方面增加噪声数据,以提升模型的鲁棒性^[9]。通过对选取的样

本集进行翻转、图像增强、添加噪声等数据增强操作^[10],以扩大样本量,使4类样本数量相对均匀。数据增强结果如图3所示。

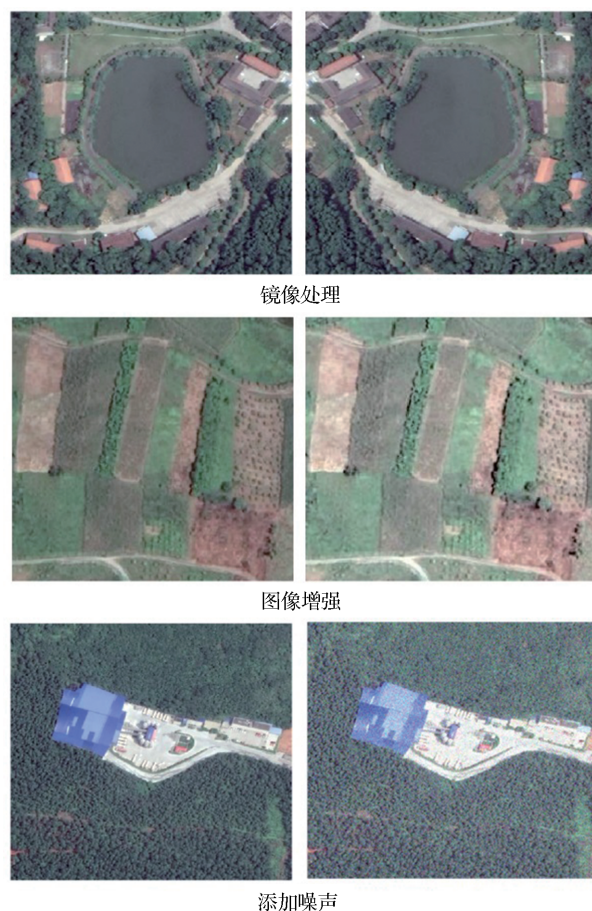


图3 数据增强技术

Fig.3 Data enhancement techniques

最后在高分辨率的遥感影像中获取建设用地、耕地、林地和水域图片,并通过数据增强手段得到每类图片各4 000张,为了防止模型的过拟合,再将各类图片按照7:2:1的比例随机分为训练、验证集和测试集,其中又对训练集进行K-fold交叉验证^[11],将样本集随机分为10份,其中7份作为训练集(train),两份作为验证集(val),一份作为测试集(test),故最终每类随机抽取2 800张放入训练集(train),800张放入验证集(val),剩余400张放入测试集(test)。

2.3 研究方法

2.3.1 VGGNet模型

VGGNet^[12]是由牛津大学的Simonyan和Zisserman提出的卷积神经网络模型,在AlexNet网络的基础上发展而来,其主要贡献在于将 3×3 的小型卷积核和 2×2 的最大池化层进行反复堆叠,并将神经

网络的深度被增加到了 16~19 层^[13]。VGGNet16 有 16 层,其中有 10 层卷积层,4 层池化层和 2 层全连接层。为了提高模型的泛化能力,本文在每个全连接层后面都增加了 Dropout 策略^[14],同时为了权衡训练时间和泛化能力最优,还引入了早停法来寻找给定泛化错误下的最小训练时间,通过设置每 10 次 Epoch 计算一次验证集的误差,将停止的标准设置为当泛化错误在连续 4 个周期内均未降低时停止模型训练。VGGNet 网络结构如图 4 所示。

2.3.2 ResNet 模型

ResNet^[15] (Residual Neural Network) 模型由前微软研究院的何凯明、张翔宇、任少卿、孙剑 4 人在 2015 年共同提出。该模型提出通过运用残差块 (Residual Blocks) 对深层次的神经网络进行训练,解决了传统深度卷积神经网络在反向传播的过程中容易造成梯度消失,导致训练模型效果不佳的问题。

2.3.3 DenseNet 模型

稠密卷积神经网络 (Dense Convolutional Network, DenseNet)^[16] 中的所有层 (具有匹配的要素图大小) 直接相互连接,保证了网络中各层之间的最

大信息流。通过每一层都接受它前面所有层的输出以保持前馈性质,并将其自己的特征映射传递给所有后续层。DenseNet 的整体结构主要由两部分构成:一是与特征图具有相同尺寸的稠密块 (Dense Blocks);二是过渡块 (Transition Layers)。上述模型在很多相关论文中都有所涉及,本文不再赘述。

2.3.4 Mask R-CNN 模型

Mask R-CNN 作为一个目前广泛应用的、高效的对象实例分割框架,能够对输入数据中的目标对象进行有效检测,同时为每个实例生成高质量的分割掩码。该方法通过添加一个分支来预测每个感兴趣区域 (ROI) 上的分割掩码,与现有分支并行进行分类和边界框回归,得到了更高效的 R-CNN^[17],结构如图 5 所示。

Mask R-CNN 在进行目标检测时对物体边缘的检测能够精确到像素级,比单任务检测具有更良好的性能,但在对掩膜及标定框进行显示时,它自身默认规定了一个阈值,当检测的精度大于该阈值时,标定框才会被显示,故最后显示标定框时有很多检测精度小于阈值的区域未被检测出来。

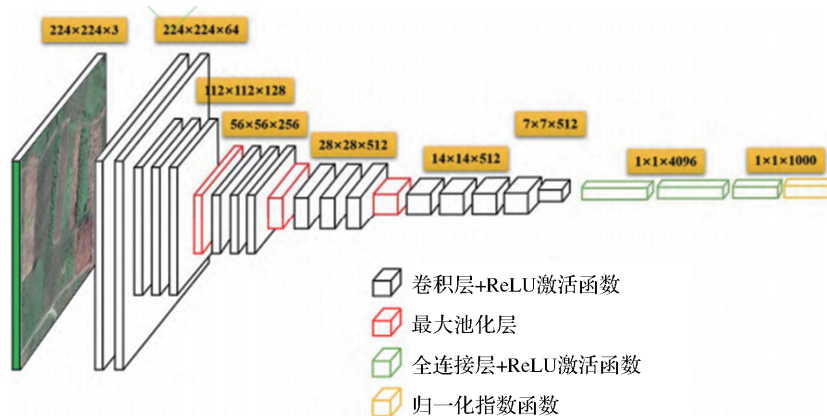


图 4 VGGNet 网络结构示意图

Fig.4 Schematic diagram of VGGNet network structure

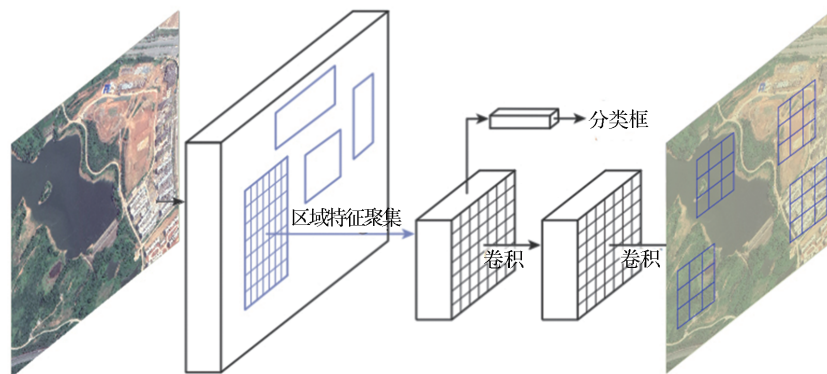


图 5 用于实例分割的 Mask R-CNN 框架^[18]

Fig.5 Mask R-CNN framework for instance segmentation

因此,本文通过改变 Mask R-CNN 的阈值,使更多的目标区域被检测到,同时为了保证被检出区域的检测准确性,并对检出区域进行裁剪,用训练过程中训练精度较好的模型的 VGG16 模型对裁剪出的区域进行类别判定,当分类结果显示该区域为建设用地时对该区域的标定框进行显示,判断流程如图 6 所示。

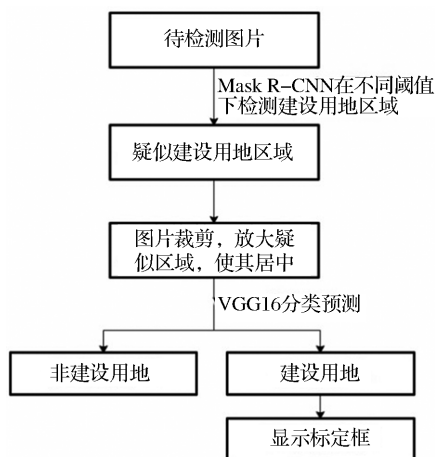


图6 VGG16+Mask R-CNN 目标检测流程

Fig.6 VGG16+Mask R-CNN target detection process

2.4 实验环境

实验环境采用 Win10 操作系统,在科学计算集成 Python 发行版 Anaconda4.2 环境下,使用 Keras 和 Tensorflow 开源框架及 CUDA-GPU 加速方案,显卡采用的 NVIDIA GeForce GTX 1080 (8G 显存) 进行 GPU 加速,其他主要硬件为 Intel(R) Core (TM) i7-6700K CPU @ 4.00GHz 处理器、64G 内存、250G 固态硬盘等。

3 结果与分析

3.1 基于卷积神经网络模型的高分辨率遥感影像分类

由训练及验证数据集的准确率及偏差变化可看出,VGG16 网络模型分别在第 2 次迭代时取得最佳结果,即此时对应模型的验证集拥有较小的偏差和较大的准确率,并通过测试集验证得到对应分辨率下的分类精度为 97.44%。ResNet 网络模型分别第 6 次迭代时得到最佳结果,此时的验证集拥有较小的偏差和较大的准确率,并通过测试集验证得到对应分辨率下的分类精度为 93.75%。DenseNet 模型分别第 25 次训练时取得最佳结果,即此时的验证集拥有较小的偏差和较大的准确率,并通过测试集验证得到对应分辨率下的分类精度为 95.13%。

VGG16 模型在前几轮训练过程中得出的模型

就能达到较好的拟合效果,随着训练次数的增加,训练集偏差不断降低,而验证集偏差开始上升,发生过拟合。相比于 VGG16,ResNet50 模型在训练过程中的精度和偏差都有较大波动,且训练时间相对较长,这是由于 ResNet50 具有更深的网络层数和更多的参数,所提取的特征更复杂和抽象,在进行分类时考虑更多的细节和特征值。在分类精度方面,ResNet50 对林地与水域的区分效果普遍不好,且测试精度显著低于 VGG16。DenseNet 模型在训练过程中验证集偏差及精度有较大波动,但总体结果优于 ResNet50 模型,稍逊于 VGG16。从影像分类结果的统计表格中可以看出 VGG16 模型耗时相对较少,迭代次数最少,精度最高;ResNet50 模型耗时最少,但精度最低;DenseNet 模型耗时与迭代次数都是最多的,但 DenseNet 模型精度要比 ResNet50 模型要高。不同模型训练过程中数据集准确率和偏差如图 7,分类结果如表 1 所示。

3.2 基于深度学习模型的高分辨率遥感影像建设用地目标检测

3.2.1 基于 Mask R-CNN 的建设用地目标检测模型训练与结果分析

实验运用 Mask R-CNN 模型(阈值=0.5),以上文中建立的遥感影像数据的建设用地类的样本集为例,对样本集中的建设用地进行目标检测。通过在 3 741 张建设用地遥感影像瓦片数据中挑选出了 3 000 张图片作为训练 Mask R-CNN 模型的样本集,图像中多为建设用地与其他几种土地利用类型同时出现的情况,故需要人工在样本集中对建设用地区域进行一一标注。本次实验选用深度学习专用的图像标注工具 VGG Image Annotator 对 3 000 张样本进行标注,对标注好的数据集按 8:1:1 随机划分为训练集、验证集和测试集,将样本集输入模型进行训练,并用测试集测试模型精度,最终得出的结果如表 2 所示。

从表 2 的结果可以看出,单独使用 Mask R-CNN 模型对样本进行训练,得出的训练结果准确率较低,且标定框的精度也不高。结合用于模型训练的样本集对实验结果进行分析,在对样本集进行人工标注的过程中将建设用地划分的较为分散,一个样本中往往包含多个检测目标,这在一定程度上加大了模型的复杂度,导致测试准确率偏低。在 300 张测试数据中一共包括 865 个检测目标,但单独运用 Mask R-CNN 模型只检测出了其中的 564 个目

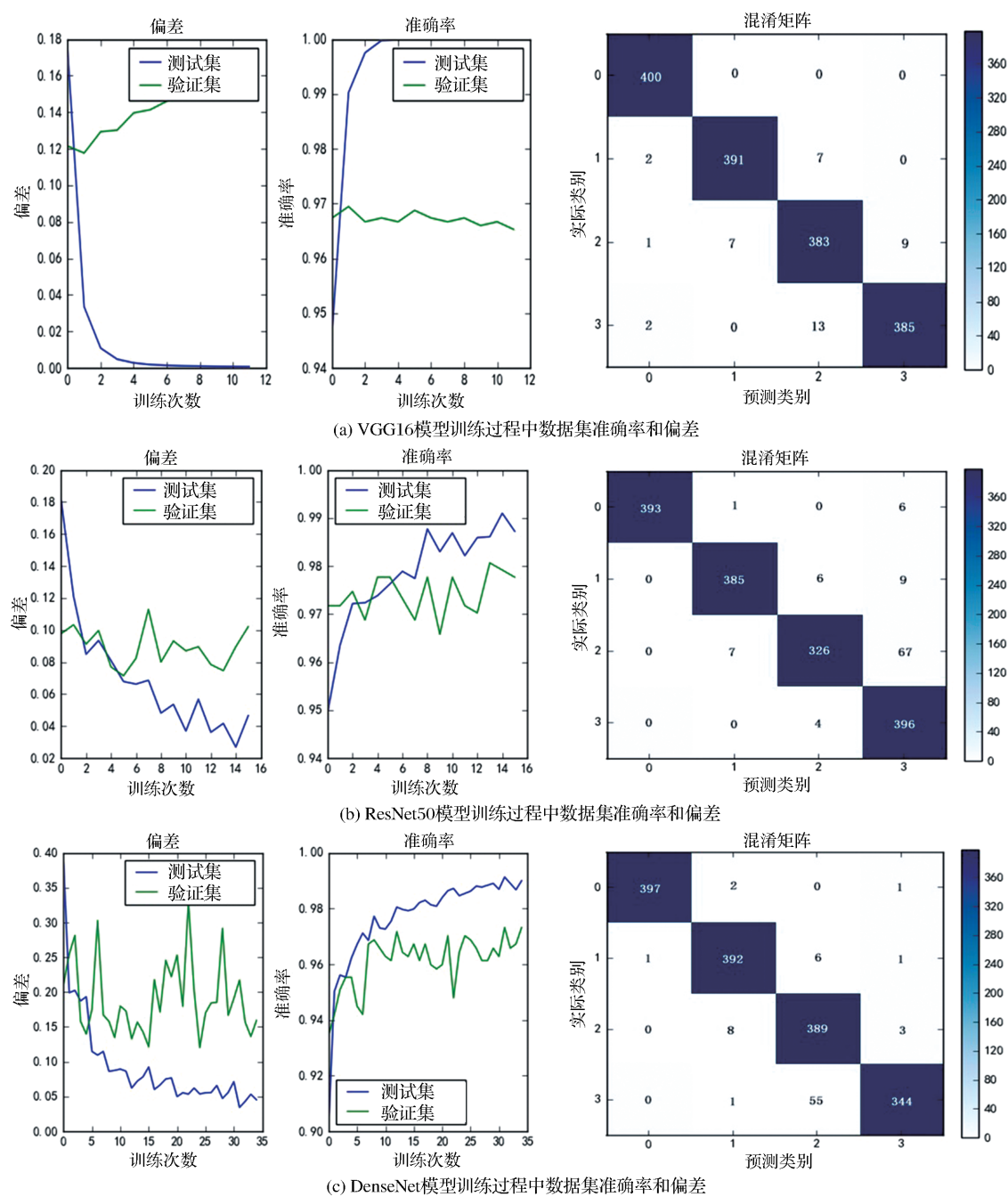


图 7 基于不同模型的高分辨率遥感影像分类

Fig.7 Classification of high-resolution remote sensing images based on different models

表 1 不同模型对高分辨影像分类结果

Table 1 Classification results of high-resolution images by different models

模型	VGGNet	ResNet	DenseNet
训练时间/s	900	640	1 440
测试集精度/%	97.44	93.75	95.13
迭代次数/次	12	16	35

标,漏检率高达 34.8%,且对于边界的划分也不够精确,标定框精度只有 76.34%,模型效果还不能有效地对建设用地进行目标检测,急需对模型进行优化

表 2 Mask R-CNN 训练结果

Table 2 Results of Mask R-CNN training

	标定框 精度/%	测试集 数量	测试集 目标数	准确 目标数	准确率 /%
Mask R-CNN	76.34	300	865	564	65.20

以满足实际需求。

3.2.2 结合 VGG16 和 Mask R-CNN 的建设用地目标检测方法优化

针对上文直接运用 Mask R-CNN 对建设用地

进行目标检测容易发生漏检,导致准确率较低的问题,本实验采用 VGG16 与 Mask R-CNN 相结合的方式对其进行优化。Mask R-CNN 在进行目标检测时对物体边缘的检测能够精确到像素级,比单任务检测具有更好的性能,但在对掩膜及标定框进行显示时,它自身默认规定了一个阈值,当检测的精度大于该阈值时,标定框才会被显示,故最后显示标定框时有很多检测精度小于阈值的目标区域未被检测出来。对 Mask R-CNN 模型设置不同的阈值,使更多的目标区域被检测到,同时为了保证被检出区域的检测准确性,并对检出区域进行裁剪,用训练过程中精度较好的模型的 VGG16 模型对裁剪出的区域进行类别判定,当分类结果显示该区域为建设用地时对该区域的标定框进行显示。测试集数量为 300,测试集目标数为 865,其测试结果对比如表 3 和图 8 所示。

表 3 不同阈值下 Mask R-CNN 训练结果
Table 3 Mask R-CNN training results under different thresholds

Mask R-CNN 阈值	标定框精度/%	准确率/%
0.1	85.35	97.46
0.2	84.32	95.61
0.3	87.93	94.34
0.4	79.69	72.32
0.5	78.54	65.2
0.6	69.84	49.94
0.7	66.32	46.59
0.8	64.87	44.74
0.9	61.56	38.84

当阈值设为 0.5 时,对比单独使用 Mask R-CNN,VGG16 与 Mask R-CNN 联合模型的标定框精度要高。

对设置不同阈值的 VGG16+Mask R-CNN 的组合模型测试集精度进行比较,随着阈值的提高,被检测出的目标从 843 减少到 336,准确率降低了接近 60%。通过比较设置不同阈值的优化模型与原模型对测试集中建设用地目标检测结果的准确率和标定框精度,评估模型的效果,选择最佳模型组合和阈值。当阈值设置为 0.3 时,标定框的精度最高,达到 87.93%,且具有较高的准确率,达到了 94.34%。

综合分析实验结果,当阈值设置为 0.3 时,标定框精度最高,同时具有较高的准确率,当降低阈值时,准确率会有小幅提升,但考虑到程序的运行时

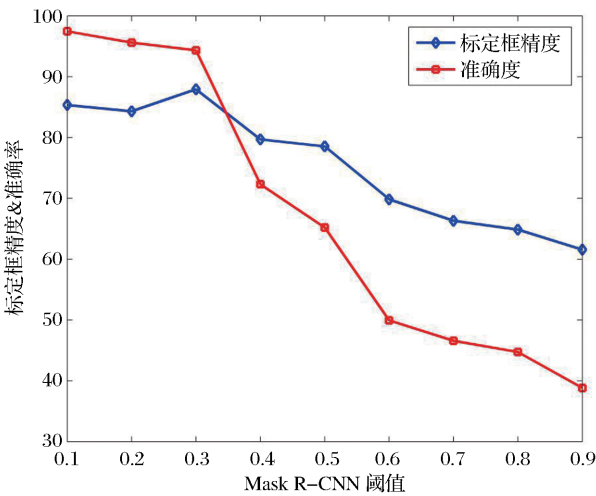


图 8 不同阈值下 Mask R-CNN 训练结果
Fig.8 Mask R-CNN training results under different thresholds

间,所以本文将阈值选取为 0.3 得到最佳效果,将其与只单独使用 Mask R-CNN(阈值=0.5)对图像进行建设用地目标检测,效果如图 9 所示。



图 9 Mask R-CNN 与 VGG16+Mask R-CNN 处理效果
Fig.9 Mask R-CNN and VGG16+Mask R-CNN

通过上图对比发现,通过选择合适阈值后的 VGG16 与 Mask R-CNN 相结合算法与单独使用 Mask R-CNN 相比,在一定程度上避免了漏检和精度不准的问题,并且改进后的算法在检测建设用地方面表现出了更强的适应性和鲁棒性,改进后的算法将原始图片划分为了更精细的多个检测区域,保持 Mask R-CNN 优良特性的同时,提高了目标区域的检出率和准确率。但由于在对样本集进行人工标记建设用地边界时并没有一个确定严格的标准,具有一定的主观性,故优化后的模型在对提取边界的处理上仍然存在目标区域与背景区域做出明确划分的问题。

4 结 语

遥感影像的分类及目标检测是遥感领域研究

的基本问题,通过对高分辨率遥感影像中的多种土地利用类型进行分类及目标检测,能够大大提高在对土地利用类型进行审计时的工作效率和检测精度。如何准确且高效地对遥感影像中的多种土地利用类型进行分类检测也是一直以来的一个研究热点,本文研究了利用 Mask R-CNN 模型结合 VGG16 分类模型对高分辨率遥感影像下的土地利用类型进行分类和建设用地目标检测,对遥感影像下的多种土地利用类型实现了高效且准确地分类和检测。主要结论如下:①遥感影像样本集的建立:使用翻转、旋转、平移等数据增强方法,既增加了样本集的数据量,又使训练的模型有更好的泛化能力,能够更好地进行后续辅助目标检测。②基于卷积神经网络模型的高分辨率遥感影像土地利用类型分类:VGG16 对高分辨遥感影像的分类精度要优于 DenseNet 模型和 ResNet50 模型,且训练所需时间相对较短,迭代次数相对较少。DenseNet 模型在训练所需时间和迭代次数方面相比于 ResNet50 模型要差,但 DenseNet 模型的精度要比 ResNet50 模型要高。③结合 Mask R-CNN 与 VGG16 实现了遥感影像中建设用地的目标检测:当 Mask R-CNN 模型阈值设置为 0.3 时,优化后的模型得到最好的检测效果,相比于单一使用 Mask R-CNN 模型,既能对目标检测有较高的准确率又较好地控制了软件运行的时间,对于多目标的检测和边界的提取精度上有了较大提高,得到了更具有现实意义的目标检测结果。

尽管本文利用 Mask R-CNN 与 VGG16 相结合的优化算法对高分辨率遥感影像中建设用地的目标检测进行了研究,与传统的 Mask R-CNN 目标检测方法相比取得了更好的实验结果,但在许多方面还需要进一步深入的研究。①在结合 Mask R-CNN 与 VGG16 进行算法优化的过程中,Mask R-CNN 模型阈值的确认方法是通过对其设置不同的值,再比较模型最后的精度等指标进行最优阈值的选取,最终得出的最佳阈值实际上只是最优值的一个近似值,后续需要通过选取更合理的阈值确定方法寻找最佳阈值的精确值。②本研究只是对 4 种典型的土地利用类型进行了分类,以及对建设用地一类目标进行了目标检测,后续研究将针对更多的自然类型进行分类和目标检测。

参考文献(References):

- [1] Hu Wei, Gao Bochuan, Huang Zhenhang, *et al.* Tree-structured Convolutional Neural Network Optimization for Semantic Segmentation of Remote Sensing Images in Urban Areas [J]. Chinese Journal of Graphics, 2020, 25(5): 197-206. [胡伟, 高博川, 黄振航, 等. 树形结构卷积神经网络优化的城区遥感图像语义分割[J]. 中国图象图形学报, 2020, 25(5): 197-206.]
- [2] Deng Zhipeng, Sun Hao, Lei Lin, *et al.* High-resolution Remote Sensing Image Target Detection based on Multi-scale Deformation Feature Convolution Network [J]. Journal of Surveying and Mapping, 2018, 47(9): 1216-1227. [邓志鹏, 孙浩, 雷琳, 等. 基于多尺度形变特征卷积网络的高分辨率遥感影像目标检测[J]. 测绘学报, 2018, 47(9): 1216-1227]
- [3] Ding Xiaohui. Classification of Hyperspectral Remote Sensing Images based on ant Colony Algorithm [D]. Changchun: Graduate School of Chinese Academy of Sciences (Northeast Institute of Geography and Agricultural Ecology), 2016. [丁小辉. 基于蚁群算法的高光谱遥感影像地物分类研究[D]. 长春: 中国科学院研究生院(东北地理与农业生态研究所), 2016.]
- [4] Pan Yifan, Zhang Xianfeng, Yu Hongfeng, *et al.* Urban Land Cover Classification based on Texture Information of Joint Kuaizhou-1 Image [J]. Remote Sensing Technology and Application, 2016, 31(1): 194-202. [潘一凡, 张显峰, 于泓峰, 等. 联合快舟一号影像纹理信息的城市土地覆盖分类[J]. 遥感技术与应用, 2016, 31(1): 194-202.]
- [5] Xianxia. High-resolution Remote Sensing Image Scene Classification based on Multi-feature Deep Learning [D]. Kunming: Yunnan Normal University, 2019. [鲜霞. 基于多特征深度学习的高分辨率遥感影像场景分类[D]. 昆明: 云南师范大学, 2019.]
- [6] Cao Zhaowei. Classification of Island Remote Sensing Images and Features based on Artificial Neural Network [D]. Shanghai: Shanghai Ocean University, 2016. [曹兆伟. 基于人工神经网络的海岛遥感影像地物分类研究[D]. 上海: 上海海洋大学, 2016.]
- [7] Ali S S, Raza I U R, Kazmi S S A, *et al.* Comparison of Pixel-based and Object-based Classification for Glacier Change Detection [C] // Second International Workshop on Earth Observation & Remote Sensing Applications. IEEE, 2012.
- [8] Luo Zhiyong, Li Xiaodong. Research on High Performance Tile Map Service based on Database Storage Scheme [J]. Geography and Geographic Information Science, 2013, 29(3): 52-55, 112. [罗智勇, 黎小东. 基于数据库存储方案的高性能瓦片地图服务研究[J]. 地理与地理信息科学, 2013, 29(3): 52-55, 112.]
- [9] Li Xin, Tang Wenli, Yang Bo. Semantic Segmentation of High Resolution Remote Sensing Image Using Depth Residual Network [J]. Journal of Applied Sciences, 2019, 37(2): 282-290. [李欣, 唐文莉, 杨博. 利用深度残差网络的高分遥感影像语义分割[J]. 应用科学学报, 2019, 37(2): 282-290.]
- [10] Yu Ximing, Hong Shuo, Yu Jinxiang, *et al.* Research on the Enhancement Method of Visible Light Remote Sensing Image Ship Target Data [J]. Journal of Instrumentation, 2020, 41(11): 261-269. [于希明, 洪硕, 于金祥, 等. 可见光遥感图像船舶目标数据增强方法研究[J]. 仪器仪表学报, 2020, 41(11): 261-269.]

- [11] Xu Chenyu. Research and Improvement of YOLOv3-based Hazardous Goods Detection Model [D]. Lanzhou: Lanzhou University, 2019.[徐晨玉. 基于YOLOv3的危险物品检测模型研究与改进[D]. 兰州:兰州大学, 2019.]
- [12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-scale Image Recognition[C]// The 3rd International Conference on Learning Representations (ICLR2015).
- [13] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks [C] // European Conference on Computer Vision. Springer, Cham, 2014:818-833.
- [14] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. The Journal of Machine Learning Research, 2014, 15 (1) : 1929-1958.
- [15] He K M, Zhang X Y, Ren S Q, *et al.* Deep Residual Learning for Image Recognition [J]. IEEE Computer Society, 2016:770-778.doi:10.1109/CVPR.2016.90.
- [16] Huang G, Liu Z, Van der Maaten L, *et al.* Densely Connected Convolutional Networks[S/OL]. <http://arxiv.org/abs/1608.06993>, 2017.doi:10.1109/CVPR.2017.243.
- [17] Ren S Q, He K, Ross Girshick, *et al.* Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6) : 1137-1149. doi: 10.1109/TPAMI.2016.2577031.
- [18] He K, Gkioxari G, Dollar P, *et al.* Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018;1-1.doi:10.1109/TPAMI.2018.2844175.

High Resolution Remote Sensing Image Construction Land Detection Combined with VGGNet and Mask R-CNN

Chen Min^{1,2}, Pan Jiawei², Li Jiangjie², Xu Lu², Liu Jiamin^{1,2}, Han Jian³, Chen Yiyun^{2,4}

(1.Guangzhou Urban Planning Survey & Design Survey Research Institute, Guangzhou 510060, China;

2.School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China;

3.Guangxi Huayao Space Information Technology co. LTD, Nanning 530031, China;

4.State Key Laboratory of Soil and Sustainable Agriculture, Chinese Academy of Sciences, Nanjing 210008, China)

Abstract: To address the problem that most current deep learning models can only discriminate land use types for cropped images of high-resolution remote sensing images, this paper combines VGGNet and Mask R-CNN to carry out a study on intelligent construction land target detection. On the basis of establishing remote sensing image datasets of four types of land use types in the study area, we compare the classification accuracy of three convolutional neural network models, VGGNet, ResNet and DenseNet, and select the neural network model with the best classification effect, VGGNet and Mask R-CNN, to achieve intelligent construction land target detection. The results show that: (1) the classification accuracies of the three convolutional neural network models VGGNet, ResNet and DenseNet are 97.44%, 93.75% and 95.13%, respectively, and the VGG16 model has the least number of iterations and relatively less training time; (2) the Mask R-CNN threshold setting has an important influence on the target detection accuracy, when the threshold is set to is 0.3, the joint model of VGG16 combined with Mask R-CNN has the highest calibration frame accuracy for construction land detection. Also the joint model has higher accuracy than the single use of Mask R-CNN model for construction land detection, and shows more adaptability and robustness.

Key words: Convolutional Neural Network(CNN); Target detection; Image classification; High resolution remote sensing image; Construction land