

# 基于卷积神经网络的无人机遥感影像 农村建筑物目标检测

任媛媛<sup>1,2</sup>, 张显峰<sup>2,3</sup>, 马永建<sup>1,2</sup>, 杨启原<sup>1,2</sup>, 汪传建<sup>1,2</sup>, 戴建国<sup>1,2</sup>, 赵庆展<sup>1,2</sup>

(1.石河子大学信息科学与技术学院,新疆 石河子 832000)

(2.国家遥感中心新疆兵团分部,新疆 石河子 832000)

(3.北京大学遥感与地理信息系统研究所,北京 100871)

**[摘要]** 将深度学习应用于遥感影像目标识别,提出基于卷积神经网络的无人机遥感影像农村建筑物的目标检测方法,用端到端的方式训练 Faster R-CNN 网络模型,并应用于农村建筑物的快速精确识别.该方法包括基于 RPN 网络的区域建议和基于 Inception v2 的卷积神经网络模型训练.为了训练和测试模型,通过无人机采集南疆地区的农村建筑物遥感影像,并人工标注建立了农村建筑物的数据集,在 TensorFlow 深度学习框架上通过对该数据集目标检测验证了模型.结果表明,基于改进的卷积神经网络目标检测方法对无人机遥感影像进行快速准确识别的总体精度超过 90%,通过初始参数更新,模型收敛更快,对无人机遥感影像地物分类和目标识别具有一定的参考意义.

**[关键词]** 建筑物,检测,无人机,深度学习,卷积神经网络,Faster R-CNN

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2019)03-0029-08

## Target Detection of Rural Buildings in UAV Remote Sensing Images Based on Convolutional Neural Network

Ren Yuanyuan<sup>1,2</sup>, Zhang Xianfeng<sup>2,3</sup>, Ma Yongjian<sup>1,2</sup>, Yang Qiyan<sup>1,2</sup>,  
Wang Chuanjian<sup>1,2</sup>, Dai Jianguo<sup>1,2</sup>, Zhao Qingzhan<sup>1,2</sup>

(1.College of Information Science and Technology, Shihezi University, Shihezi 832000, China)

(2.Xinjiang Production and Construction Corps Division of National Remote Sensing Center of China, Shihezi 832000, China)

(3.Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China)

**Abstract:** With deep learning applied to object recognition of remote sensing images, a method of object detection for rural buildings based on convolution neural network is proposed in the paper. The improved Faster R-CNN network model is trained in an end-to-end way and applied to rural buildings with the rapid and accurate identification. Specifically, the method mainly includes region recommendation based on RPN network and convolutional neural network model training based on Inception v2. In order to train and test the improved model, the remote sensing images of rural buildings in southern Xinjiang Region are collected by UAV, and the data set of rural buildings is established by manual labeling. Finally, the model is validated by the object detection of the data set with TensorFlow deep learning framework. Experimental results show that the overall accuracy of fast and accurate recognition of UAV remote sensing images based on the improved convolution neural network object detection method exceeds 90%. By updating the initial parameters, the model converges faster, which has a certain reference value for the classification and object recognition of UAV remote sensing images.

**Key words:** buildings, detection, UAV, deep learning, convolutional neural network, Faster R-CNN

农村建筑物识别是观察农村的土地变化和经济发展的重要手段.近年来,高分辨率遥感影像开始覆盖人们生活的各个方面,不再只针对发达的城市地区.农村建筑物多为农村居民点的房屋,识别这些居民点用地,对农村土地利用情况进行研究,则是新农村建设的基础<sup>[1]</sup>.中国作为传统的农业大国,农村发展

收稿日期:2019-07-05.

基金项目:国家重点研发计划(2017YFB0504203)、国家自然科学基金(41461088)、兵团科技计划(2016AB001、2017DB005).

通讯联系人:汪传建,博士,教授,研究方向:机器学习、数据挖掘. E-mail:wcj\_inf@shzu.edu.cn

至关重要,快速、准确地获得农村建筑物的信息对于统筹城乡发展和节约集约用地有着重大的帮助. 新疆作为一个面积辽阔、多民族交融的地区,农村建筑物信息的快速获取又被赋予了更多的意义. 通过对识别出的建筑物进行研究,可以了解农村人口的变化情况,从而维护公共安全,促进新疆地区社会的长治久安和平稳定发展.

随着高分辨率对地观测系统(earth observing system,EOS)和自动目标识别系统(automatic target recognition,ATR)的发展,高分辨率遥感影像的目标识别也有了很大的进步. 通过目标识别获得的目标种类、位置等信息,在无人驾驶、军事战争、公共安全、交通监管等方面起到了重要的作用. 因此,基于高分辨率遥感影像的目标识别研究具有重要的意义<sup>[2-3]</sup>. 同时,遥感影像时间和空间分辨率的提高,也为灾情现势的及时评估及应急和减灾提供了帮助<sup>[4]</sup>. 随着遥感技术的发展,遥感影像的数量与分辨率都有了飞速提升,基于这些高质量的遥感影像数据,使得农村建筑物的识别与变化检测成为了可能. 而无人机遥感则是以灵动性见长,能采集到人力所采集不到的信息. 无人机遥感的时间周期短,能弥补卫星因为天气、时间等原因无法采集遥感影像的缺陷. 在传统农业、国情监测、物流配送等行业中起着巨大的作用<sup>[5]</sup>.

目标识别作为计算机视觉领域的热点,推动了遥感影像地物识别、分类的进步. 在目前的通用目标检测方法中,不仅包含了 Haar 特征<sup>[6]</sup>+Adaboost 算法、Hog 特征+SVN 算法<sup>[7]</sup>和 DPM 算法<sup>[8]</sup>等传统算法,同时也出现了基于深度学习的新型目标检测方法,该方法主要分为 two-stage 和 one-stage 两种检测算法.

2013 年 Sermanet P 等<sup>[9]</sup>提出 Overfeat 算法中使用图像缩放和滑动窗口的方法,结合 bounding box 回归,使得卷积神经网络具备了定位能力,可用于解决目标检测问题. 2014 年 Girshick R 等提出了基于区域的卷积神经网络(regions with CNN,RCNN),通过产生与类别无关的区域推荐,使用卷积神经网络进行特征提取,最后使用指定类别的线性 SVM 来对候选区域进行分类<sup>[10]</sup>,但该网络要求图像输入大小固定. 为解决该问题,2015 年 He K M 等提出了 SPPNet,使得网络输入不必固定大小,通过在全连接层上加入一层空间金字塔池化(spatial pyramid polling,SPP)产生固定大小的特征,使得网络可以处理任意尺寸和比例的图像<sup>[11]</sup>,但该方法在处理性能上稍显不足. 同年,Girshick R 提出了 Fast R-CNN 网络,借助多任务损失函数使得训练中可直接更新所有层的参数,提高了网络性能<sup>[12]</sup>. 2016 年 Ren S Q 等人提出了速度更快的 Faster R-CNN,通过用区域建议网络(region proposal network,RPN)代替原先的候选区域算法,使得提出的候选区域更为精简、准确,同时可共享卷积层参数,减少了计算量,使得该网络在同等算法中检测速度更快<sup>[13]</sup>. 以上算法都将目标检测问题分为产生候选区域和对候选区域分类两个阶段,属于 two-stage 检测算法. 而 2016 年 Redmon J 等提出的 YOLO 算法通过一个神经网络直接得到边界框以及其所属的类别概率,将目标检测变为回归问题. 由于检测过程使用单一网络,因此可直接进行端到端(end-to-end)的检测性能优化,大幅提高了检测速度<sup>[14]</sup>,但对小目标检测结果较差. 2016 年 Liu W 等提出的 SSD 网络增加了多尺度预测,改善了小目标检测的不稳定性,精度也相对提高<sup>[15]</sup>. 这些 one-stage 检测算法不需要产生候选区域的步骤,直接产生物体的类别分数和位置. 考虑到遥感影像中农村建筑物目标成片或成点出现以及对精度的需求,本文选择以 Faster R-CNN 网络为基础进行农村建筑物的目标检测.

近年来,深度学习理论也被尝试应用于建筑物或农村建筑物识别上. 曲景影<sup>[16]</sup>等通过卷积展开技术改进 LetNet-5 网络,提高了高分辨率遥感影像目标识别的精度和效率. 陈文康<sup>[17]</sup>使用改进的 CaffeNet 对农村地区建筑物遥感影像进行分类,分类精度为 95%. Audebert N<sup>[18]</sup>等使用深度全卷积网络进行建筑物检测,以不同尺度卷积核的卷积神经网络分别预测不同尺寸的建筑物,均取得了较好的效果. Saito S<sup>[19]</sup>等将卷积神经网络原有的 softmax 分类用通道抑制多分类(channel-wise inhibited softmax)进行代替,实现了遥感影像中建筑物的自动提取. 金永涛<sup>[20]</sup>等使用面向对象与深度学习的方法对遥感典型地物进行提取,解决了典型地物混分问题. 刘文涛<sup>[21]</sup>提出一种级联式全卷积神经网络,并在深度卷积神经网络的设计中使用了特征复用和特征增强,实现建筑物的自动精确提取. 上述研究虽均取得了较好的识别效果,但对建筑物位置的精准检测研究还相对较少.

基于此,本文提出一种基于卷积神经网络的遥感影像农村建筑物目标检测算法,该算法引入 Inception v2 结构对 Faster R-CNN 网络进行改进,共享卷积层参数,减小网络参数规模,提高网络的性能,在 TensorFlow 深度学习框架上进行试验,调整参数并与其他网络模型对比分析. 试验结果表明,改进后的 Faster R-CNN 网络对农村建筑物的识别更快速准确,具有一定的优势.

## 1 基于 Faster R-CNN 的目标检测

Faster R-CNN 通过对 R-CNN、Fast R-CNN 进行优化,使用 RPN 代替以往的 Selective Search 算法<sup>[22]</sup>来构造候选区域,减少候选框生成数量,实现了更高的目标检测精度和更快的检测速度。如图 1 所示,该方法包括两部分:(1)区域建议的深度全卷积网络,(2)Fast-RCNN<sup>[12]</sup>。

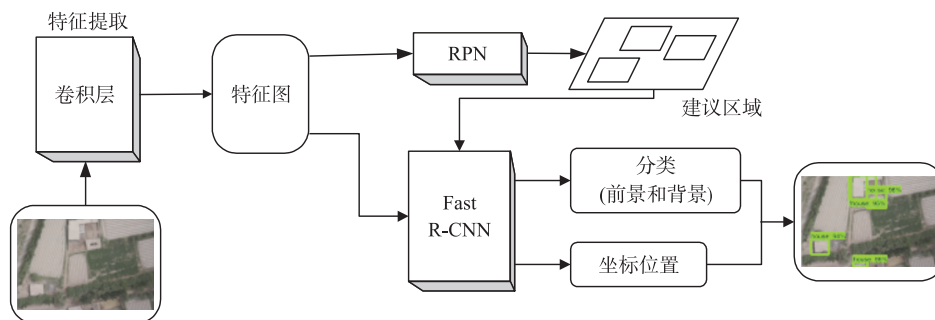


图 1 Faster R-CNN 网络结构

Fig. 1 Faster R-CNN network structure

### 1.1 区域建议网络 (RPN)

RPN 是一个以平移不变性作为基础的全卷积网络。其基本组成部分(卷积、池化和激励函数)作用在局部输入域,只依赖相对空间坐标<sup>[23]</sup>,即目标物体经过平移之后,仍然具有相应的特征。在目标检测问题中,具有平移不变性很重要,当目标物体位置变换后,依然具有其特征才能被准确检测出来。以任意大小的图像作为 RPN 的输入,输出一组目标候选框,针对每个候选框都有一个目标得分,且与 Fast R-CNN 目标检测网络共享一个卷积层。RPN 是目前较为先进的候选区域提取方法,每个图像提出的候选框数量从 Fast R-CNN 的 2 000 个减少到 300 个,减少了计算时间,提高了网络效率。同时,该网络可预测每个位置的目标边界和对应分数。经过端到端的训练,RPN 可生成高质量的候选框。由于和目标检测网络共享一组卷积层,RPN 在共享卷积层的最后输出卷积特征图上滑动一个小网络,输入窗口为  $3 \times 3$ ,映射到一个低维特征,以 VGG-16 为例,由于最后输出卷积特征图为 512 维,则此处的低维特征即为 512 维,此处特征连接到两个全连接层,分别用来边界框回归(定位)和边界框分类(分类)。

受益于 anchor 机制,RPN 可以监测各种尺度及不同长宽比的目标物体。anchor 机制的尺度为  $128^2$ 、 $256^2$ 、 $512^2$ ,长宽比为  $2:1$ 、 $1:1$ 、 $1:2$ 。对应每个滑动窗口,可组合成 9 种不同尺度比例。例如,针对每个滑动窗口预测  $k$  个区域建议,则边界框回归层具有  $4k$  个输出,包括这  $k$  个矩形建议框的坐标点。而边界框分类层的输出则是  $2k$ ,即是否为目标物体的概率。在本文中,由于组合出 9 种尺度比例的区域建议框,则  $k=9$ 。在训练 RPN 的过程中,需要划分正负样本,有两种情况可以标记为正样本:①对于每个推荐区域,计算其与实际真实边界框的交并比 IoU,具有最高值的 anchor 标记为正样本;②IoU 值大于 0.7 的样本标记为正样本。这样保证了每个区域起码对应一个正样本。而 IoU 值低于 0.3 的 anchor 被标记为负样本,若一个 anchor 既未被标记为负样本,也未被标记为正样本,则其在训练中是无用的。

在训练 RPN 的过程中由于负样本占大多数,若对所有的 anchor 进行优化,会使其偏向于负样本。故在每张图像中随机选取 256 个 anchor,其中正负样本比为  $1:1$ ,若正样本少于 128 个,则使用负样本填充,保证有足够 256 个 anchor 用于训练。

### 1.2 卷积神经网络

本文在训练 Fast R-CNN 网络时,采用占用资源更少、训练速度更快的端到端的训练方式(end-to-end)。Faster R-CNN 提供 3 种训练方式:Inception、ResNet、NasNet,对应不同大小网络模型。本文选择 GoogleNet Inception v2<sup>[24]</sup>作为训练的网络结构。

Inception 系列模型为了应对类似 VGG 网络那样在一定深度就性能饱和的问题,以及降低内存与计算的开销,减少了使用参数,使其更具有实用价值。不同于常见的卷积神经网络,GoogleNet 增加了层宽,每个卷积层卷积核数目的增加有助于模型处理局部信息,增强其表示能力,也使模型收敛得更快。为减少内部协变量转移(internal covariate shift)问题(即在训练神经网络的过程中,每层的输入都会因为前一层参数的变化而改

变,使得难以构造一个饱和和非线性的模型),Inception v2<sup>[25]</sup>模型加入了批规范化(batch normalization, BN)层,使每一层的输出规范化.此外,如图 2 所示,参照 VGG 网络结构,用两个较小的 3×3 的卷积核替代了原先 5×5 的卷积核.在保持感受野的情况下,减少了参数量,增强了网络的非线性表达能力,如图 3 所示.

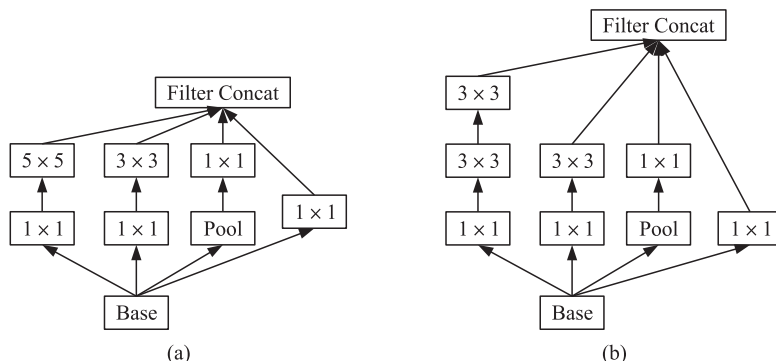


图 2 改进型 Inception 网络结构

Fig. 2 Improved Inception network structure

RPN 和 Fast R-CNN 共享特征,二者交替训练.首先训练 RPN,再参照以上改进训练 Fast R-CNN,然后由 Fast R-CNN 的网络参数初始化 RPN,并重复该过程.该网络结构层次深,通道数多,提取到的特征更明显,有利于进行目标检测.

## 2 遥感影像目标检测试验

### 2.1 数据预处理

本文试验所使用的农村建筑物高分辨率遥感影像数据是 2018 年 5 月 5 日于新疆兵团第三师图木舒克市 51 团通过纵横 CW-20 复合翼无人机平台获取

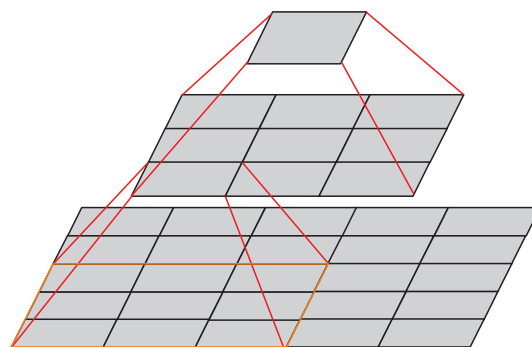


图 3 感受野变化

Fig. 3 Changes in receptive field

的.经预处理得到图像尺寸为 7 360×4 912,共计 386 幅遥感影像,其可见光影像分辨率达 0.05 m.由于遥感影像尺寸较大,且图像中目标较为稀疏,背景较多,考虑到试验时正负样本的比例,对其进行适当剪裁,并人工挑选,尽量使数据集中图片中含有目标物体,且目标物体尽量不在分割线上.对这些遥感影像进行剪裁,归一化为 500×334 像素的图像,获得数据集共计 1 400 张图片,训练集与测试集比例为 4:1,使用 LabelImg 软件对数据进行人工标注.

参照 VOC2012 的数据集格式,所使用的数据分为 3 部分:①ImageSets/Main/保存了具体数据集的索引,其中 test.txt 包括了测试集图片的索引,train.txt 包括了训练集的所有图片索引,trainval.txt 则是两者的集合;②Annotations 保存了标签数据,xml 格式文件包括了物体的标签、图片的尺寸、物体的 bounding box 坐标(包括四角的坐标等);③JPEGImages 保存了图片内容. LabelImg 软件生成的就是 Annotations 中的 xml 文件.

### 2.2 数据格式转换

由于本文试验在 TensorFlow 深度学习框架上进行,需要输入专门的 TFRecord Format. TFRecord 格式文件使用了“Protocol Buffer”数据描述语言,将结构化数据序列化类似于 XML,以二进制数据编码存储的 TFRecord 文件只占用一个内存块,内存一次加载一个二进制文件即可,不需要像 VOC2012 数据集一样零散存储,对大型网络训练更为友好,故将 VOC2012 数据集格式转变为 TFRecord 数据格式.

TFRecord 格式对标签进行统一存储.本文通过两个脚本将 VOC2012 格式数据集转化为 TFRecord 格式文件.先将 Annotations 中的 xml 格式文件统一使用分隔符方式记录为 csv 格式文件,每栏名称分别为‘filename’、‘width’、‘height’、‘class’、‘xmin’、‘ymin’、‘xmax’和‘ymax’,该脚本名称为 xml\_to\_csv.py.再将 csv 格式文件处理为 TFRecord 格式,在 Example 中写入对应数据.对于训练集和测试集分别运行上述代码,得到 train.record 和 test.record 文件.



### 2.3 数据增强

在卷积神经网络训练过程中,数据依赖性大,对数据要求高.在样本数量不够多,或某类样本数量少、特征不足的情况下,增强数据有助于网络训练,防止模型过拟合.考虑到采用无人机一次采集的数据质量较为稳定,只采用平移和旋转方式对数据进行增强<sup>[3]</sup>.

在数据预处理阶段已对遥感数据进行裁剪,由于目标物体可以出现在数据集的任何位置,这种平移(也即从原有图像上裁剪一部分)可避免对同一目标同一位置的过拟合.旋转处理则是考虑到真实情况下,目标物体会以各种角度出现在图像中,考虑到试验需要定位到物体具体坐标的操作,以及增加模型的泛化能力,本文以 $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ 这3种角度对图像进行旋转,使得 bounding box 的边界与坐标轴平行.

### 2.4 试验结果及分析

本文在 TensorFlow 深度学习框架上,以配置了 Intel i7-7700K CPU 处理器的机器对 1 400 张无人机遥感影像图片(分辨率为  $500 \times 334$ )进行训练.图片进行人工标注等处理后,将标定的信息与图片以 TFRecord 形式输入 Faster R-CNN 开始训练.选择 Inception v2 网络模型作为训练网络,迭代次数 15 000 次,初始学习率 0.000 2,以端到端的方式来训练 RPN 和卷积神经网络,得到农村建筑物目标检测模型.由于硬件的限制,选择 batch 大小为 1.学习率大小的不同设置对于模型的收敛速率及模型的效果有较大的影响.学习率设置过大会导致损失函数 loss 无法收敛,而设置过小则会导致陷入局部最优解,影响模型精度.本文采取逐渐增大学习率的方法,当 loss 趋向平稳时,则增大学习率,如表 1 所示.

在训练中对比了有无初始参数赋值的情况.以 TensorFlow model zoo 中的 faster\_rcnn\_inception\_v2 模型对本文的模型进行初始化,模型收敛很快,且初始 loss 值较低,模型识别率也较好.无初始化的模型参数初始值随机,一开始 loss 值较高且收敛慢,同为 5 000 次训练时效果没有进行初始化的模型好.部分结果如图 4 所示,其中(a)是初始化过的,(b)是未进行初始化的,可以看出(a)效果较好,而(b)将覆盖了地膜的土地错误地划分成房子.

表 1 学习率变化设置

Table 1 Learning rate change settings

迭代次数/次	学习率	loss
0	0.000 20	2.114 1
500	0.000 22	1.187 4
1 500	0.000 24	1.021 2
2 500	0.000 26	0.876 7
3 500	0.000 28	0.701 4
4 500	0.000 30	0.653 7
7 000	0.000 35	0.499 1

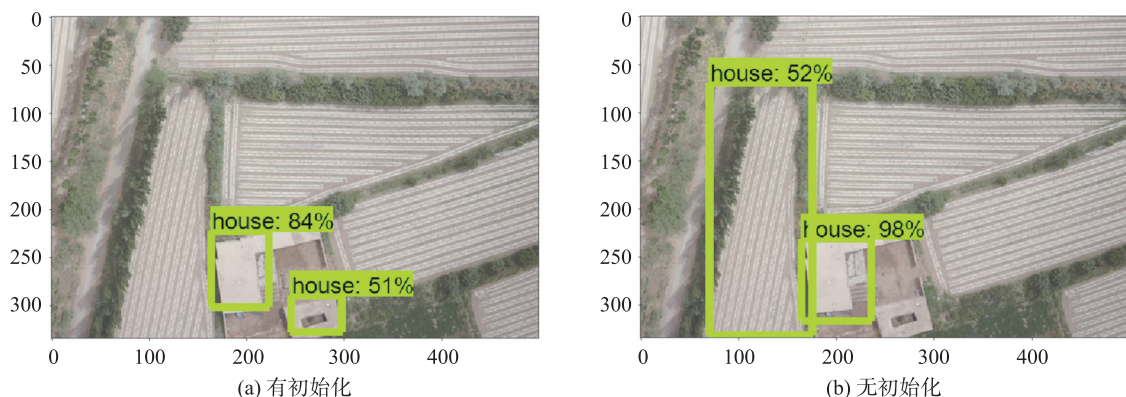


图 4 有无初始化操作的结果对比

Fig. 4 Comparison of initialization results with or without initialization operation

如表 2 所示,本文算法对于农村建筑物有较好的检测效果.算法对于单张尺寸大小为  $500 \times 334$  的输入图像在无 GPU 加速的情况下,检测时间约为 1.41 s.而在同样情况下使用 Faster R-CNN+ResNet101 网络,单张图片所耗时间过长,相同时间无法达到同样的迭代次数,但精度相对较高.在有 GPU 加速的情况下,可考虑使用 ResNet 应对更复杂的情况.同等情况下,SSD+MobileNet v1 处理训练速度最快,但识别精度却不是很高.由于遥感图像中负样本较多,且目标较为稀疏,大小不一,故存在检测的边界框不能正好包围目标的情况.同时,人工标记时也会存在一定误差,导致边界框不能正好包围目标,因而 IoU 值较低,但仍然能正确估计目标位置,以上统计数据是在 IoU 阈值为 0.1 的情况下进行的.本文同时对比最小距离法(Minimum Distance)、最大似然法(Maximum Likelihood)和支持向量机(Support Vector Machine)等传统机器学习方法,其中最大似然法和支持向量机的试验效果如图 5 所示.可以看出,地膜部分在最大似然法

中被归为房屋,出现了虚警的情况. 即使 SVM 识别精度较高,也依然出现了虚警. 其中,SVM 识别精度达到了 84.2%,但训练时间及耗费的资源较多. 综上,本文所使用的网络识别精度可达 90.5%,在精度和效率上都具有较大优势.

表 2 模型测试性能对比

Table 2 Performance comparison of model testing

深度学习模型	识别精度/%	识别耗时/s
Faster R-CNN+inception v2(进行了初始化)	90.5	1.41
Faster R-CNN+inception v2(未进行初始化)	78.5	1.45
Faster R-CNN+ResNet101	85.2	4.20
SSD+MobileNet v1	67.8	0.75
Minimum Distance	82.4	—
Maximum Likelihood	83.1	—
SVM	84.2	—

此外,本文算法在识别小目标上较不稳定,如图 6 所示,小目标存在漏检的现象,识别精确度也较差. 图 6(a)是模型在训练 4 000 次时的效果,(b)是模型在训练 5 000 次时的效果,可以发现对小目标检测较不稳定. 其主要原因是卷积神经网络随着深度越深,浅层特征信息丢失越严重,故而小目标的信息在池化过程中更易丢失. 其次,IoU 阈值对于小目标影响较大,在参数更新中,小边界框的误差不易缩小. 最后,在人工标注的数据集中,小目标易标注不准确,易误标. 未来可以考虑增大图像分辨率,或改进网络结构,如结合高层次的语义信息和低层次的位置信息来进行目标检测,以提高小目标的检测效果.

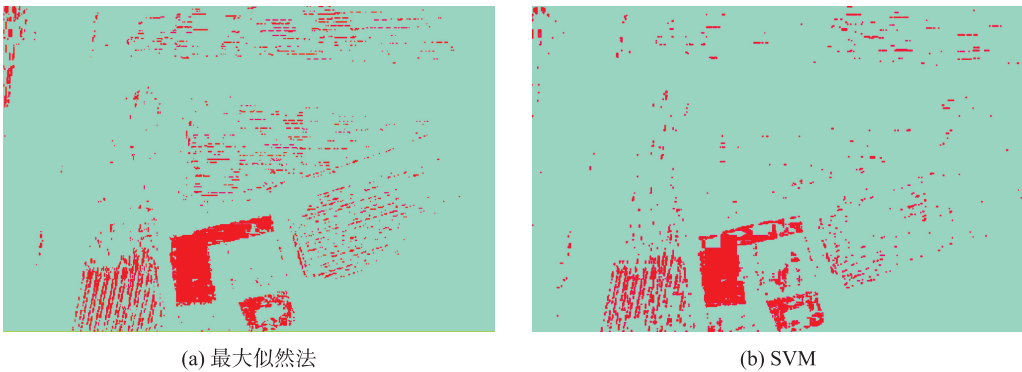


图 5 传统方法检测结果

Fig. 5 Test results with traditional methods



图 6 目标检测结果

Fig. 6 Target detection results

3 结论

本文在 TensorFlow 深度学习框架上搭建了 Faster R-CNN 的深度学习模型,卷积神经网络部分采用了 inception v2 网络,整理并标注了一个南疆图木舒克周边的农村建筑物无人机遥感影像数据集. 通过识别高分辨率无人机遥感影像中的农村建筑物,分析了不同参数和卷积模型对于目标识别的影响. 对比了基

于 ResNet101 的 Faster R-CNN 网络和 SSD+mobilenet v1 网络, Faster R-CNN+inception v2 在基于高分辨率无人机遥感影像的目标识别问题上, 识别精度和速度最为均衡, 取得了良好的检测效果. 而算法针对小尺寸目标识别精度较低的问题未来将进一步改进.

## [参考文献] (References)

- [1] 郑辛煜. 农村居民点遥感信息提取与时空演变规律研究[D]. 杭州:浙江大学, 2017.  
ZHENG X Y. Research on rural residential areas remote sensing information extraction and spatial-temporal evolution[D]. Hangzhou: Zhejiang University, 2017. (in Chinese)
- [2] 刘扬, 付征叶, 郑逢斌. 高分辨率遥感影像目标分类与识别研究进展[J]. 地球信息科学学报, 2015, 17(9): 1080–1091.  
LIU Y, FU Z Y, ZHENG F B. Advances in target classification and recognition of high resolution remote sensing images[J]. Geo-information science, 2015, 17(9): 1080–1091. (in Chinese)
- [3] 王金传, 谭喜成, 王召海, 等. 基于 Faster R-CNN 深度网络的遥感影像目标识别方法研究[J]. 地球信息科学学报, 2018, 20(10): 1500–1508.  
WANG J Z, TAN X C, WANG Z H, et al. Faster R-CNN deep learning network based object recognition of remote sensing image[J]. Geo-information science, 2018, 20(10): 1500–1508. (in Chinese)
- [4] 李小文. 汶川震灾中遥感的应急与反思[J]. 遥感学报, 2008, 12(6): 838.  
LI X W. Emergency response and reflection of remote sensing in Wenchuan earthquake disaster[J]. Journal of remote sensing, 2008, 12(6): 838. (in Chinese)
- [5] 李德仁, 李明. 无人机遥感系统的研究进展与应用前景[J]. 武汉大学学报(信息科学版), 2014, 39(5): 505–513.  
LI D R, LI M. Research progress and application prospect of UAV remote sensing system[J]. Geomatics and information science of Wuhan university, 2014, 39(5): 505–513. (in Chinese)
- [6] VIOLA P A, JONES M J. Rapid object detection using a boosted cascade of simple features[C]//Computer Society Conference on Computer Vision & Pattern Recognition. Kauai, USA: IEEE, 2001.
- [7] SCHULDT C, LAPTEV I, CAPUTO B. Recognizing human actions: a local SVM approach[C]//Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK: IEEE, 2004.
- [8] FELZENSZWALB P F, MCALLESTER D A, RAMANAN D. A discriminatively trained, multiscale, deformable part model[C]//Conference on Computer Vision & Pattern Recognition. Anchorage, USA: IEEE, 2008.
- [9] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: integrated recognition, localization and detection using convolutional networks[C]//International Conference on Learning Representations. Banff, Canada: ICLR, 2014.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 580–587.
- [11] HE K, ZHANG X, REN S, 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.
- [12] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1440–1448.
- [13] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. Montreal, Canada: NIPS, 2015: 91–99.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 779–788.
- [15] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector[C]//European Conference on Computer Vision. Amsterdam, Netherlands: ECCV, 2016: 21–37.
- [16] 曲景影, 孙显, 高鑫. 基于 CNN 模型的高分辨率遥感图像目标识别[J]. 国外电子测量技术, 2016(8): 45–50.  
QU J Y, SUN X, GAO X. High resolution remote sensing image target recognition based on CNN Model[J]. Foreign electronic measurement technology, 2016(8): 45–50. (in Chinese)
- [17] 陈文康. 基于深度学习的农村建筑物遥感影像检测[J]. 测绘, 2016, 39(5): 227–230.  
CHEN W K. Remote sensing image detection of rural buildings based on deep learning[J]. Surveying and mapping, 2016,

- 39(5):227–230.(in Chinese)
- [18] 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. Taipei, China; AFCV, 2016:180–196.
- [19] SAITO S, YAMASHITA Y, AOKI Y. Multiple object extraction from aerial imagery with convolutional neural networks[J]. Electronic imaging, 2016, 60(1):10402–1/10402–9.
- [20] 金永涛, 杨秀峰, 高涛, 等. 基于面向对象与深度学习的典型地物提取[J]. 国土资源遥感, 2018, 30(1):22–29.  
JIN Y T, YANG X F, GAO T, et al. Typical ground object extraction based on object-oriented and deep learning[J]. Remote sensing for land & resources, 2018, 30(1):22–29.(in Chinese)
- [21] 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. Object detection in high resolution remote sensing images based on fully convolution networks[J]. Geo-information science, 2018, 20(11):1562–1570.(in Chinese)
- [22] UIJLINGS J R R, SANDE V D, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2):154–171.
- [23] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE, 2015:3431–3440.
- [24] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. Lille, France; ICML, 2015.
- [25] SZEGEDY C, VANHOUCKE V, LOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE, 2016:2818–2826.

[责任编辑:严海琳]