基于卷积神经网络面向自然场景建筑物识别技术的移动应用

许博鸣[1] 刘晓峰[1] 周京正[2] 张福全[1] 业巧林[1]

[1]（南京林业大学信息科学技术学院 南京 210037）

[2]（中华人民共和国公安部科技信息化局 北京 100741）

（liuxiaofeng@njfu.edu.cn）

A Convolutional Neural Network based Mobile Application for Identification of Buildings in Natural Scene

Xu Boming[1], Liu Xiaofeng[1], Zhou Jingzheng[2], Zhang Fuquan[1], and Ye Qiaolin[1]

[1]（*College of Computer Science and Technology, Nanjing Forestry University, Nanjing* 210018）

[2]（*Bureau of Science and Information Technology, Ministry of Public Security of the People's Republic of China, Beijing* 100741）

Abstract Due to the presence of background noise in natural scenes and the interference of complex factors such as illumination, rotation, and shooting angle, it is very difficult to identify the image of buildings in natural scenes. Aiming at the dependence of traditional building extraction methods on human design and the improvement of building edge feature extraction algorithm, based on the image recognition technology using convolutional neural network to classify landmark buildings in natural scenes and the realistic demand to transplant CNN models to mobile terminals, the real-time identification of complex scenes requires real-time recognition of the MobileNet bottleneck layer through Keras, and the addition of a new classifier for transfer learning. A large number of data augmentation and test set augmentation are applied to the input image. After two stages of transfer learning, The accuracy of 96.5% 93.9% and 94.4% was achieved within 480 iterations in three test set. Compared with other feature extraction algorithms, CNN has the advantages of non-transformation and automatic extraction of features, achieves higher accuracy in a shorter period of time. At the same time, mobilenet weighs only occupy 15.3 MB with high precision and less calculation, which can be widely transplanted to mobile devices. The system based on model transplantation has the functions of photo recognition, photo album recognition, menu display, etc., providing mobile platform users with a convenient and simple tool to obtain the information of buildings in natural scenes quickly and accurately.

Key words transfer learning; deep learning; convolutional neural network; mobile system transplantation

摘要 由于自然场景中背景噪声的存在,以及光照、旋转、拍摄角度等复杂因素的干扰,使得自然场景中对建筑物的图像识别难度较大.针对传统建筑物提取方法对人为设计的依赖,以及对建筑物边缘特征提取算法的改进,基于卷积神经网络（Convolutional Neural Networks,CNN）对自然场景中地标建筑物进行分类的图像识别技术,以及将 CNN 模型移植到移动端实现复杂场景的实时识别的现实需求,通过 Keras 框架获取 MobileNet 瓶颈层后加入新的分类器进行迁移学习,对输入图片进行大量的图像增强技术和测试集增强技术,经过三个阶段的迁移学习 480 次迭代后在三个测试集上分别达到 98.3%、93.9%、94.4%的准确率.相对比其他的特征提取算法,CNN 具有平移不变形以及自动提取特征等优点,在较短的时间内获得较高准确率的同时,MobileNet 的权

基金项目：国家自然科学基金项目（61871444 , 31670554）
This work is supported by the National Natural Science Foundation of China (61871444 , 31670554)

通信作者：刘晓峰（liuxiaofeng@njfu.edu.cn）

重仅有 15.3MB,兼顾计算量和精度,可以广泛移植到移动端设备.基于模型移植的移动端系统兼具拍照识别、相册识别、菜单展示等功能,为移动平台用户提供一个方便简捷的工具来快速准确地判断自然场景中建筑物的信息.

关键词 迁移学习;深度学习;卷积神经网络;移动平台移植

中图法分类号 TP391

建筑物作为城市地区中重要的地理标识,也是人类活动区域的重要特征,在复杂自然场景中对其的精确识别可以为城市规划、旅游业等提供重要支撑.而基于移动设备对自然场景中的图像进行识别已经成为深度学习领域重要的的热点之一,随着图像识别技术近年来的不断发展,使得在移动端对高分辨率的自然场景图像进行精确分类成为现实.

传统从高分辨率图像识别建筑物主要有两种形式:一是通过高分辨率遥感影像[1,2]中建筑物的自动提取.Qi-Ming Qin 等[3]在理论上提出了基于小波描述的建筑物识别方法并实验证明了该方法的可行性.Ghandour Ali 等[4]使用阴影验证的建筑物检测技术对高分辨率卫星图像实现自动提取,达到了超过 95%的精确度.二是通过自然场景中的建筑物的自动提取.Wang Liying 等[5]使用基于体素的 3D 建筑物检测算法对 ISPRS 城市数据集达到了 96.11%的完整性和 95.87%精确度.Yongkwon Kim 等[6]使用搜索范围确定和边缘跟踪技术对移动车辆摄像机的捕获视频进行建筑物识别,达到了 88.9%的准确率.

自 2012 年自 Alexnet[7]模型的提出,深度学习[8]中的神经网络[9]快速发展,卷积神经网络[10]被广泛应用到图像识别处理技术中.曲延云等[11]利用 Sobel 算子进行水平与垂直边缘的检测,并使用支持向量机实现了对自然建筑物图像的检测.Shahzad, Muhammad 等[12]在极高分辨率的合成孔径雷达图像中使用深度全卷积神经网络,实现了 93.84%的平均像素准确率.Kang Zhao 等[13]提出了一种将 Mask R-CNN 与建筑物边界正则化相结合的方法,更好的提取给定卫星图像中定位的建筑物多边形.

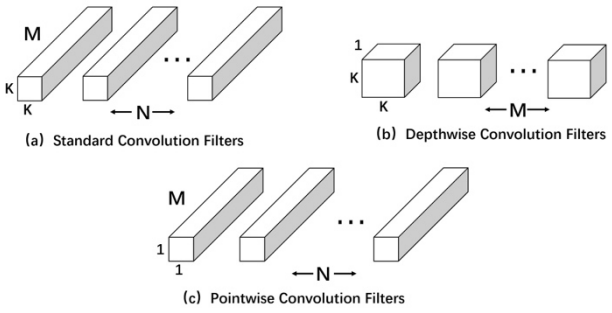
本文基于卷积神经网络对自然场景建筑物进行分类的思想,并考虑到移植至移动平台进行实时识别的重要性,利用 MobileNet[14]作为基础模型进行多种版本的迁移学习并给出训练原理,不仅可以在减少网络参数的同时获得较高的识别率,也让移动端快速识别成为了现实.

本文第 1 节给出了模型的介绍,建立以及详细的训练过程,包含迁移学习得以奏效的原因和训练细节.第 2 节阐述了训练集的收集以及预处理,包含图像增强技术和测试集增强技术.第 3 节给出了将 CNN 模型移植到移动端并开发出完整识别系统的过程和细节.第 4 节给出了实验数据对比,验证了多种迁移学习版本的有效性和测试集增强后对模型鲁棒性的分析.最后总结全文,表明了该学习方法的高效性以及移动系统的完整性.

1 建筑物识别模型

1.1 模型介绍

MobileNet[14]是 google 于 2017 年提出的一种小巧而高效的神经网络模型.它基于流线型架构,使用深度可分离卷积来构建轻量级神经网络模型.它将标准卷积分解成深度卷积和点卷积（如图 1 所示）,并在每个卷积层后加入了批标准化[15]操作来避免反向传播算法[16]中梯度消失的现象.其中 K 为卷积核的长度和宽度,M 为前一层卷积的通道数量,N 为卷积核个数.另外 MobileNet 模型引入了两个超参数宽度乘数 α 和分辨率乘数 β,前者可以以一定的比例缩减整个网络的通道数量,后者可以改变特征图的输入尺寸.本文使用精度最高的 1.0 MobileNet-224 模型.



Figue 1. The architecture of MobileNet: standard convolution filters(a) are replaced by depthwise convolution filters(b) and pointwise convolution filtets(c)

图 1. MobileNet 模型：将标准卷积核分解成深度卷积核和点卷积核。

1.2 模型建立

通过 Keras 框架获得 MobileNet 的瓶颈层,随后使用卷积核大小为 7×7、步长为 1 的均值池化,输出结果为 1×1×1024,展开(Flatten)后相当于一个 1024 结点的全连接层,使用 dropout[17]来防止过度拟合.为了加快收敛速度,再连接一个 512 结点的全连接层并使用 dropout,使用 relu 作为激活函数,最后连接一个 26 个结点的网络层对应建筑物种类,使用 softmax 作为激活函数.drpoout 参数均为 0.4.模型建立过程如下

```
1.  mobile=MobileNet(include_top=False,weight=
    'imagenet',input_shape=[224, 224, 3])
2.  last = mobile.output
3.  x=AveragePooling2D((7,7),strides=(1,1),
    padding='valid')(last)
4.  x = Flatten()(x)
5.  x = Dropout(0.5)(x)
6.  x = Dense(512, activation='relu')(x)
7.  x = Dropout(0.5)(x)
8.  x=Dense(26,activation='softmax')(x)
9.  model= Model(inputs=mobile.input,outputs=x)
```

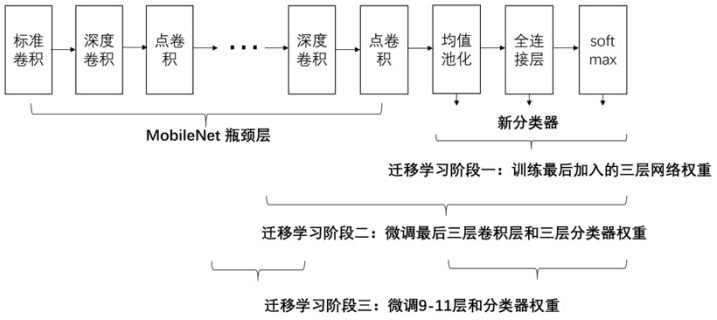
1.3 迁移学习训练细节

随着神经网络的快速发展,在图像分类领域中迁移学习[18]开始变得尤为重要.对于 imagenet 这样庞大的数据集,已经有许多人使用若干块 GPU 训练数周的时间,实现了对 1000 个物种的分类.这些训练好的权重虽然不能识别其他领域的图片,但确对图像有着很强的特征提取能力.从 ZFnet[19]的可视化反卷积中可以清晰的看出过滤器在每一层学习了何种特征,并且随着网络层数的不断增加,高层过滤器提取的特征更加复杂.

根据 Yosinski J[18]所述的迁移学习过程可以看出,低层过滤器学习到的特征具有共性,可以很好的迁移到另一个网络中.而在网络的中层表现出的性能很差,这是由于它们处于十分关键位置,负责把低层特征转换成具有针对性的高层特征,这些相邻层之间的参数存在着很高的耦合性,他们往往与很强的依赖性(fragile co-adapted features),通过协同学习得到.到了网络的最后几层,由于模型的特征已经基本构建完成,这时问题已经从一个复杂的高维非线性问题变成了相对低维的线性问题,重新学习的内容越来越少.故本文将迁移学习的重点放在训练中层网络的特征转换上,并且在该阶段学习时必须将整个中层网络打开,而不能只打开某一层,否则会破坏相互之间的依赖性.另外也训练了网络高层来微调复杂特征提取.由于获取的 MobileNet 瓶颈层为 14 层(深度卷积和点卷积视为一层),根据比例我们将 9-11 层定义为 MobileNet 的中层网络,这一部分重点调整,而将 12-14 层定义为网络高层,这一阶段进行微调.

故本文的迁移学习分为三个阶段.阶段一中将瓶颈层之前的权重冻结,只训练池化层,全连接层和 softmax 层.在阶段二中,微调高层,即 12-14 层和分类器.阶段三开放 9-11 层和最后的三层分类器.阶段一由于需要有较快的收敛速度,故使用 Adam 优化器,并设置学习速率为 0.0001.阶段二和阶段三所以使用带动量的随即梯度下降算法(SGD),参数设置为学习速率 0.01, 学习速率衰减为 1e-6, 动量参数为 0.9.

整个迁移学习阶段的过程如图 2 所示



Figue 2. MobileNet model in transfer learning

图 2. MobileNet 迁移学习过程

2. 训练集收集与预处理

2.1 数据集收集

本文从散落的数据集中共收集 1548 张来自世界各地共 26 种著名建筑物作为地标识别,其中 319 张图片作为测试集,测试集中包含 205 张从视频中截取的图片,这些图片更符合自然场景中的要求.剩下的 1229 张图片大约按照 3:1 的比例,916 张作为训练集,313 张作为验证集.

2.2 图像增强技术

由于从网络上收集到的数据中的图像尺寸大小各不相同,需要对图像进行预处理.Alexnet[7]中先把图片的最短边缩减成 256,再随即裁剪出 224×224 的图片并进行水平翻转,另外对训练集图像中的 RGB 通道进行了主成分分析(PCA).VGG[20]模型在裁剪图片上使用定长裁剪和不定长裁剪.在不定长裁剪中,会把图片的最短边 S 随即裁剪到 256-512 之间,使得模型对图片有了更大的感受视野.在 Googlenet[21]中进行了多尺度采样,从原始图片裁剪出的尺寸大小从 8%到 100%

的均匀分布随机选择一个,纵横比从 3/4 和 4/3 随机选择一个.

本文利用了 Keras 中的 ImageDataGenerator 和 fit_generator 进行图像增强处理.

具体 ImageDataGenerator 图像生成器代码及参数如图所示:

```
train_datagen = ImageDataGenerator(
    width_shift_range=0.2,
    height_shift_range=0.2,
    zoom_range=0.3,
    rescale=1./255,
    fill_mode='nearest',
    horizontal_flip=True,
)
```

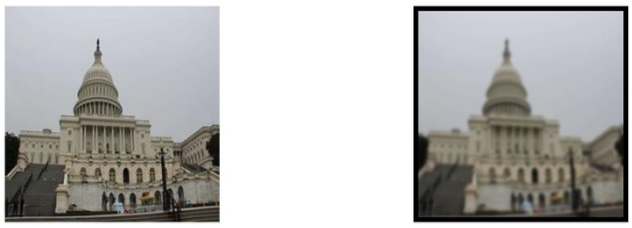
在 ImageDataGenerator 中对训练集作出了如下处理:随机水平偏移幅度为 0.2,随机竖直偏移幅度为 0.2,随机缩放幅度为 0.3,并进行水平翻转.另外把所有图片的像素值归一化到[0,1]之间(rescale=1./255),可以有效的解决反向传播算法[17] (Back Propagation,BP 算法)中梯度消失的问题,增加了学习速度,规律的图像分布也可以让模型更好的寻找全局最优解.由于考虑到实际中的建筑物识别很少存在倒立的情况,所以并未采用垂直翻转增强技术.

另外本文使用 ImageDataGenerator 中的 flow_from_directory()方法从文件路径生成增强数据,采用多线程的方式,一边生成增强后的图片一边传入内存,和传统的未经图像增强处理直接全部将图片读入内存,或将图像增强后生成的图片保存至硬盘再加载到内存相比,不用一次将所有的数据读入内存当中,大大减小内存压力.

2.3 测试集增强技术

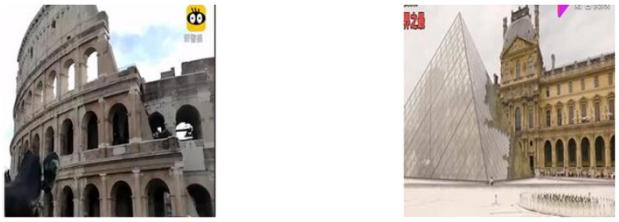
由于本文移动端采取的拍照技术是直接获取缩略图(见 3.2 节),为了更好的显示真实场景中的准确度,把原测试集进行一定程度的高斯模糊作为测试集改变后的版本一,将从视频中截取后的图像加入原测试集作为版本二,在测试模型准确率时将会在三个测试集上横向比较,最大程度还原现实自然场景的识别精度.

高斯模糊后的图像更符合移动端拍照功能获得的图像,也很好的验证了模型的鲁棒性,图 3 给出高斯模糊化后的图像.而截取视频中的图像则能更好的还原自然场景中的图像,图 4 给出截取的图像.



Figue 3. Images before and after Gaussian blur

图 3. 显示高斯模糊前后的图像。



Figue 4. Display images captured in the video

图 4. 显示截取视频中的图像。

3. 移动端移植技术

3.1 图像识别过程实现

首先将 keras 保存的.h5 模型文件转换 tensorflow 的.pb 模型文件.在 Android Studio 中加入 tensorflow 的依赖项,进行移动系统的开发.手动实现三个基本图片预测函数.

(1) argmax 函数:获取最大预测值的标签,这是由于 keras 调用函数式 API 后对应的 predict 函数返回的是一个 n 维向量对应 n 个浮点数,分别代表识别为该类标签的概率值,使用 argmax 函数后即可获取最大概率标签值.

(2) processBitmap 函数:将手机相机或相册获取图片的 bitmap 缩减成规定尺寸,根据原尺寸与化简尺寸的比例来进行缩减,本文中规定尺寸为 224×224.

(3) normalizeBitmap:标准化操作,将输入图标 bitmap 的像素值移动到区间[0,1]的范围内,由于 bitmap 的存图原理为对每一个像素点对应 GRB 三个通道分别分配 8 个位(bit),相当于一个像素点对应 24 个位,通过位运算每次移动 8 个位就可以从 bitmap 中获取像素点的具体 RGB 数值并进行标准化操作.

3.2 系统开发细节

移动端开发过程中在利用监听器实现拍照功能时,使用的是直接获取拍照的缩略图,未使用从将照片存储到 sdk 卡中再读取高清图片.这是因为在移动端拍出的高清图片直接传给网络的代价太大,若想顺利识别必须进行有效的裁剪操作.故直接获取整个图片的缩略图既可以节省操作时间,也进一步说明模型在清晰度不高的缩略图上也可以进行精确识别,说明该模

型具有良好的鲁棒性. 而设置菜单功能可以展示该系统能够识别出所有建筑物名称和缩略图标, 在点击任何一栏建筑物后会显示完整大图以及建筑物的详细介绍.

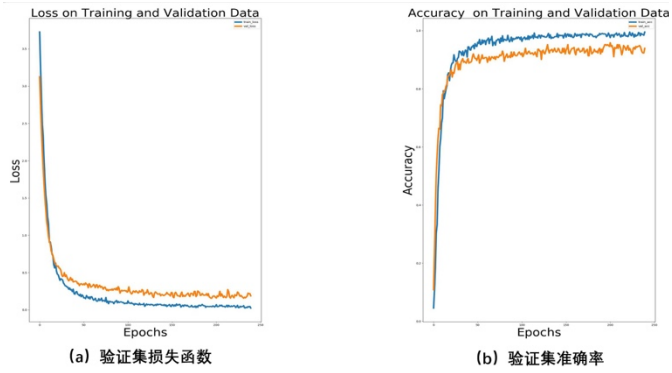
移动端界面如图 5 所示.

Figue 5. App display

图 5. 移动端展示

4. 结果分析

图 6 展示了迁移学习过程一中损失函数（图 a）和准确率（图 b）的变化情况, 横坐标为迭代次数, 蓝色曲线表示训练集, 黄色曲线表示验证集. 从图中可以看出, 由于冻结的网络层已经具有很强的特征提取能力, 所以在前 50 次迭代中, 损失函数会迅速减少, 进一步说明了迁移学习的高效性.



Figue 6. training process in transfer learning version 1

图 6. 迁移学习训练过程一

表 1 给出三个阶段的迁移学习训练过程中的训练细节,包括迭代次数、优化器、学习速率的选择以及模型准确率.可以看出微调网络高层和网络中层均可以有效提高整个模型的精确度.

Table 1. training details in different transfer learning

表 1. 不同迁移学习训练细节

	迭代次数（epoch）	准确率（%）	优化器	学习速率
迁移学习阶段一	240	93.9	Adam	0.0001
迁移学习阶段二	120	96.5	SGD+ Momentum	0.01
迁移学习阶段三	120	98.3	SGD+ Momentum	0.01

表 2 给出了迁移学习在三种数据集上的准确率.可以看出无论在使用模糊图像还是视频截取图像中,模型的准确率较最开始的清晰数据集都没有较大的下降,而整体的准确率随着迁移学习不同程度的进行有所上升,说明该模型具有较好的鲁棒性,在强化后的测试集上依然可以达到较稳定的准确率.

Table 2. three versions of training results on three test sets

表 2. 三个阶段迁移学习训练结果

	迁移学习一（训练新分类器）	迁移学习二（微调网络高层）	迁移学习三（微调网络中层）
清晰数据集	93.9%	96.5%	98.3%
模糊数据集	86.8%	91.2%	93.9%
模糊图像&视频截取图像	91.5%	94.4%	94.4%

为了体现中层网络的耦合性,本文分别训练了一同打开 9-11 层的网络权重和依次打开第 9、10、11 层的网络权重,结果如表 3 所示.可以看出一同训练中层网络可以有效的提高模型的准确度,而依次打开中层网络进行训练则会破坏结点之间的耦合性,由于中层网络充当着将低层特征转换成高层特征的重要角色,因此依赖性被破坏后导致准确度略有下降.

Table 3. Verification of the coupling of the middle layer network

表 3. 对中层网络耦合性的验证

	迁移学习阶段三（整体训练 9-11 层）	迁移学习阶段三（分开训练 9-11 层）
模糊图像&视频截取图像	94.4%	93.1%

5. 总结

本文基于卷积神经网络,使用了大量的图像增强和测试集增强技术,进行了三个阶段的迁移学习并给出详细学习过程以及原理,针对复杂自然场景中的地标建筑物实现精确分类.基于各个阶段训练后的数据对比,给出了一种高效的迁移学习训练方式，并证明了中层网络的耦合性在学习过程中的重要性.最后将模型移植到移动端实现方便简捷的识别系统,借此实现针对自然场景的实时分类.在 480 次迭代的较短时间内,达到了三个测试集 98.3%、93.9%、94.4%的识别率,同时权重仅有 15.3MB,具有保持较高识别率的同时有效降低了模型的权重空间的优点.

参 考 文 献

[1] Fan Rongshang,Chen Yang,Xu Qiheng,et al.A high-resolution remote sensing image building extraction method based on deep learning[J].Acta Geodaetica et Cartographica Sinica,2019,48(1):34-41. DOI:10.11947/j.AGCS.2019.20170638.

(范荣双,陈洋,徐启恒,王竞雪.基于深度学习的高分辨率遥感影像建筑物提取方法[J].测绘学报,2019,48(01):34-41.)

[2] Li H , Liu F , Yang S Y , et al. Remote sensing image fusion based on Deep Support Value Learning Networks[J]. Chinese Journal of Computers, 2016.

(李红, et al. 西安电子科技大学计算机学院 西安, 西安电子科技大学智能感知与图像理解教育部重点实验室、智能感知与计算国际联合研究中心、智能感知与计算国际合作联合实验室 西安, 基于深度支撑值学习网络的遥感图像融合[J]. 计算机学报, 2016, 39(8):1583-1596.)

[3] Qin Q M , Chen S J , Wang W J , et al. The building recognition of high resolution satellite remote sensing image based on wavelet analysis[C]// 2005 International Conference on Machine Learning and Cybernetics. IEEE, 2005.

[4] Ali G , Abedelkarim J . Autonomous Building Detection Using Edge Properties and Image Color Invariants[J]. Buildings, 2018, 8(5):65-.

[5] Wang L , Xu Y , Li Y . A Voxel-Based 3D Building Detection Algorithm for Airborne LIDAR Point Clouds[J]. Journal of the Indian Society of Remote Sensing, 2018.

[6] Kim Y , Lee K , Choi K , et al. Building Recognition for Augmented Reality Based Navigation System[C]// IEEE International Conference on Computer & Information Technology. IEEE, 2006.

[7] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2).

[8] Lecun Y,Bengio Y,Hinton G . Deep learning [J]. Nature,2015,521(7553):436.

[9] Li-Cheng J , Shu-Yuan Y , Fang L , et al. Seventy Years Beyond Neural Networks:Retrospect and Prospect[J]. Chinese Journal of Computers, 2016.

(焦李成, 杨淑媛, 刘芳, et al. 神经网络七十年:回顾与展望[J]. 计算机学报, 2016, 39(8):1697-1716.)

[10] Fei-Yan Z , Lin-Peng J , Jun D . Review of Convolutional Neural Network[J]. Chinese Journal of Computers, 2017.

(周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017(6).)

[11] Qu Y , Zheng N , Li C , et al. Salient building detection based on SVM[J]. Jisuanji Yanjiu yu Fazhan/Computer Research and Development, 2007, 44(1):141-147.

(曲延云, 郑南宁, 李翠华, et al. 基于支持向量机的显著性建筑物检测[J]. 计算机研究与发展.)

[12] Shahzad M , Maurer M , Fraundorfer F , et al. Buildings Detection in VHR SAR Images Using Fully Convolution Neural Networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018:1-17.

[13] Zhao K , Kang J , Jung J , et al. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018.

[14] Howard A G , Zhu M , Chen B , et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.

[15] Ioffe S , Szegedy C . Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015.

[16] Rumelhart D E . Learning Representations by Back-Propagating Errors[J]. Nature, 1986, 23.

[17] Hinton G E , Srivastava N , Krizhevsky A , et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012.

[18] Yosinski J , Clune J , Bengio Y , et al. How transferable are features in deep neural networks?[C]// Advances in Neural Information Processing Systems. MIT Press, 2014.

[19] Zeiler M D , Fergus R . Visualizing and Understanding Convolutional Networks[J]. 2013.

[20] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[21] Szegedy C , Liu W , Jia Y , et al. Going Deeper with Convolutions[J]. 2014.

Xu Boming, born in 1996. Received his BSc degree from Nanjing Forestry University, China. His main research interests include algorithm design, and machine learning.

Liu Xiaofeng, born in 1983. Lecturer, member of China Computer Federation. His main research interests include information retrieval, Web mining, and machine learning.

Zhou Jingzheng, born in 1979. Senior Engineer. Member of China Computer Federation. His main research interests include information security and intrusion detection.

Zhang Fuquan, born in 1977. Associate Professor. Member of China Computer Federation. His main research interests include computer networks and wireless networks.

Ye Qiaolin, born in 1982. Associate Professor. Member of China Computer Federation. His main research interests include machine learning and pattern recognition.

通讯作者：刘晓峰

通讯地址:南京市龙蟠路 159 号南京林业大学信息学院

邮编：210037

电话：13951071830

Email 地址：liuxiaofeng@njfu.edu.cn