

基于 FlowS-Unet 的遥感图像建筑物变化检测

顾炼¹ 许诗起¹ 竺乐庆¹

摘要 针对目前人为探察土地资源利用情况的任务繁重、办事效率低下等问题,提出了一种基于深度卷积神经网络的建筑物变化检测方法,利用高分辨率遥感图像实时检测每个区域新建与扩建的建筑物,以方便对土地资源进行有效管理.本文受超列(Hypercolumn)和 FlowNet 中的细化(Refinement)结构启发,将细化和其他改进应用到 U-Net,提出 FlowS-Unet 网络.首先对遥感图像裁剪、去噪、标注语义制作数据集,将该数据集划分为训练集和测试集,对训练集进行数据增强,并根据训练集图像的均值和方差对所有图像进行归一化;然后将训练集输入集成了多尺度交叉训练、多重损失计算、Adam 优化的全卷积神经网络 FlowS-Unet 中进行训练;最后对网络模型的预测结果进行膨胀、腐蚀以及孔洞填充等后处理得到最终的分割结果.本文以人工分割结果为参考标准进行对比测试,用 FlowS-Unet 检测得到的 F1 分数高达 0.943,明显优于 FCN 和 U-Net 的预测结果.实验结果表明,FlowS-Unet 能够实时准确地将新建与扩建的建筑物变化检测出来,并且该模型也可扩展到其他类似的图像检测问题中.

关键词 FlowS-Unet, 建筑物变化检测, 全卷积神经网络, 多尺度交叉训练, 多重损失

引用格式 顾炼, 许诗起, 竺乐庆. 基于 FlowS-Unet 的遥感图像建筑物变化检测. 自动化学报, 2020, 46(6): 1291–1300

DOI 10.16383/j.aas.c180122

Detection of Building Changes in Remote Sensing Images via FlowS-Unet

GU Lian¹ XU Shi-Qi¹ ZHU Le-Qing¹

Abstract Since manually detecting the situation of land resource utilization is arduous and inefficient, a smart building change detection method based on deep convolutional network is proposed, which can detect newly emerged or expanded buildings in each region of the high-resolution remote sensing images at real-time, thus can be used to manage the land resources efficiently. This article proposes a model named FlowS-Unet by applying refinement and other improvements to U-Net, which was inspired by hypercolumns and the refinement structure in FlowNet. First, the remote sensing images were cropped, denoised, and semantically annotated to form the dataset which is further divided into the training set and testing set, the training set is augmented to get enough training samples, and the mean value and variance of all training images are calculated and used to normalize the dataset; Second, the training set is fed into the fully convolutional network FlowS-Unet for training, which integrates multi-scale cross training, multiple losses and Adam algorithm for its optimization. Finally, the predicted result of FlowS-Unet is further post-processed with dilating, eroding and hole-filling to get the final segmentation result. By using manually segmented results as the ground truth, a comparison with several different algorithms shows that the F1 score of FlowS-Unet is as high as 0.943, which is apparently better than the predicted results of fully convolutional networks (FCN) and U-Net. Experimental results indicate that the newly emerged or expanded buildings can be accurately detected at real time with FlowS-Unet. This model can also be applied to other similar image detection problems.

Key words FlowS-Unet, change detection for buildings, fully convolutional networks (FCN), multi-scale cross training, multiple losses

Citation Gu Lian, Xu Shi-Qi, Zhu Le-Qing. Detection of building changes in remote sensing images via FlowS-Unet. *Acta Automatica Sinica*, 2020, 46(6): 1291–1300

在国土监察业务中,很重要的一项任务是监管地上建筑物的建、拆、改、扩.现阶段解决这项工作的方法主要有两种:一是依靠人工实地调研取证,但

这对于大城市来说完全靠国土局公务员全城巡查是不可能的,既耗费大量的人力、物力以及财力,又无法做到全方位实时监管国土资源利用现状;二是在各地安装高清摄像头,利用视频检测技术与 GPS 设备,建成国土资源综合动态智能监管系统.但这种方法建设成本高,时间跨度长,且只适用于小区域试点开展.

近年来,我国卫星发射次数不断上升,多颗卫星在太空运行后带来大批遥感数据,这些数据都是宝贵的历史材料,在城市发展中存在多方位的应用,如

收稿日期 2018-03-05 录用日期 2018-07-15
Manuscript received March 5, 2018; accepted July 15, 2018
浙江省自然科学基金(LY20F020002)资助
Supported by Natural Science Foundation of Zhejiang Province (LY20F020002)
本文责任编辑 刘青山
Recommended by Associate Editor LIU Qing-Shan
1. 浙江工商大学计算机与信息工程学院 杭州 310018
1. School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018

可应用于灾害监测服务、土地资源管理等领域,并推送给社会组织、政府机构甚至个人,使他们共同参与城市管理和监督.将遥感技术应用于国土资源调查时,不仅具有信息多、效率高、多层次、现实性强等特点,而且与传统方法相比较具有费用低、速度快、精度高、周期短的优势.

有研究人员采用传统图像处理方法来检测卫星图像上建筑物的变化,这些检测方法从开始的面向像元逐渐向基于对象的方法过度,提出了均质区域识别^[1]、分水岭分割^[2]、形态学房屋指数计算^[3]、能量最小化^[4]、聚类提取城市变化^[5]、先验形状约束水平集模型^[6]、自下而上/自上而下混合算法^[7]、简单几何结构法^[8]等方法和模型.然而这些方法存在以下不足:1)同一地点不同时期的图像间配准精度要求高,多数方法以图像间的精确配准为前提,但实际应用中如此高的配准精度通常难以达到,导致算法的漏检率、错检率依然很高;2)遥感数据要求严格,特定方法只适用于特定数据,如有的要求多波段影像,有的需要三维数据^[9].

深度学习技术在近几年发展迅猛,传统的深度卷积分类网络如 AlexNet^[10] 和 VGGNet^[11] 通常包含全连接层(Full connected layer),要求输入图像的大小固定.这些网络模型存在存储开销大,计算效率低以及感知区域大小受限的缺陷.针对语义分割任务,Long 等^[12] 提出的全卷积神经网络(Fully convolutional network, FCN),使得分割可以在任意大小的图像上进行,与基于分类网络的图像分类方法相比,提高了处理速度.但用 FCN 进行语义分割得到的结果仍不够精细,其低分辨率的输出牺牲了定位精度^[13].只对各个像素进行独立分类,未充分考虑像素与像素之间空间和值的关系对分类结果的影响,缺乏空间一致性. Ronneberger 等^[14] 提出的 U-Net 是对 FCN 的延伸,是一种编码器解码器结构,该网络较浅层用来解决像素定位问题,较深层用来解决像素分类问题,通过对低层次特征映射的组合,构建高层次复杂特征,来实现精确定位,解决图像分割问题,是目前扩展性较好的全卷积网络.但该网络最后一层特征是与前几层特征上采样放大后融合所得,更适应于分割所有建筑物,而对于本文任务,其特征表达能力不够,易出现过拟合现象,无法精准预测出建筑物的变化情况. Hariharan 等^[15] 针对卷积神经网络(Convolutional neural network, CNN)最后一层能较好地描述类别语义,而缺少位置等细节信息的问题提出了超列(Hypercolumn)的概念,即对应像素的网络所有节点的激活进行上采样串联作为特征,使最后一层的特征在空间上比较精细,进行目标的细粒度定位.对于物体变化检测方面, FlowNet^[16] 是将 CNN 运用于光流上检测的网

络,网络分为收缩和放大部分,其收缩部分为九层卷积结构,放大部分是将收缩部分对应的特征图经反卷积及上采样将光流预测连接起来的步骤,此步骤重复四次,每一步提升两倍分辨率.

卫星图像完全不同于分辨率高、物体差异性大、轮廓复杂、同时遮挡很严重的 ImageNet、NYUDv2 等图像数据集,卫星图像分辨率低,建筑物的差异性小、轮廓也更简单,表现为具有一定面积、长度和宽度的面状地物^[17].而且不同卫星、不同拍摄角度等复杂的遥感成像过程还会引起生成的卫星图片色调不同、建筑倾斜角度不同的问题,不同波段之间存在明显的相关性差异.文献[18–19]已经认识到针对自然图像设计的分割方法是不适用于有不同数据特征、尺寸更大的遥感图像.设计能对此类卫星遥感数据产生好的预测结果的深度学习模型是需要解决的问题之一.文献[20]使用弱监督的 SegNet 以及弱监督语义分割模型在多个国家的卫星图上进行微调检测建筑物,再将两个模型预测出的结果结合来得到更好的结果.文献[21]是对高分辨率的谷歌卫星图像从 RGB 颜色空间上转换到 Lab 色彩空间,再从 Lab 色彩空间的亮度分量上设置双阈值找出阴影,将阴影形状结合太阳照射方向自动选择出建筑物的缓冲区域来作为有监督分类的训练区域,最后对每个缓冲区域值使用改进的平行六面体监督分类器检测出建筑物.文献[22]提出先尝试找到可能的物体中心点,然后使用 R-FCN 与 ResNet-101 结合的模型根据中心点生成检测框.其他典型的方法还有三种 CNN 多尺度捕获纹理信息^[23]、两阶段 CNN 模型^[24]、结合边缘先验知识与分类得分法^[25]等.2017 年阿里云天池“广东政务数据创新大赛—智能算法赛”要求参赛者完成由不同年份的遥感图像自动检测建筑物变化情况的任务.该任务不仅要求模型能学习图像特征表达,还能学习两幅图像对应像素点之间的联系.本文作者参加了该竞赛,提出并实现了 FlowS-Unet 模型,最终取得综合第 2 名的成绩.

本文受超列和 FlowNet 中的放大部分细化结构启发,将 FlowNet 中的细化结构运用到 U-Net 并结合其他改进方法提出 FlowS-Unet 网络.超列和 FlowNet 中的细化结构相似,都是将不同层次的特征图进行若干次上采样然后进行串联作为特征,使特征在空间上更加精细,更适合于目标的细粒度定位. FlowS-Unet 模型沿用 U-Net 的 U 型结构框架,输入输出都是图像,左半部分框架不变,右半部分加入细化结构,每层都独立预测.在文献[13]中也提出,对于建筑物的提取方面,是需要不同层特征共同来确定该像素是否为建筑物的.例如,高分辨率特征图对于精确捕捉建筑物的边界是需要的,而深层低分辨率特征图可对地面是建筑物还是空地进行分类.

实验表明该模型对识别两期同一地区的 QuickBird 卫星影像差异比 FCN、U-Net 更有效, 从而检测出建筑物变化情况, 检测到的边缘更加明晰, 精确度有很大提升. 因拍摄时间不同, 卫星遥感影像中的建筑物变化信息复杂, 除水泥地上扩增建筑之外, 还有从水体、裸土、植被等面上扩增的建筑物; 同时建筑物的尺度也具有多样性, 包括了地铁站、居民区、城中村、厂房等不同形态大小的建筑; 已有的方法检测建筑物变化效果不佳, 而 FlowS-Unet 模型具有很强的学习能力, 且易于扩展, 可以在数据集较小的情况下表现出较强的泛化能力, 性能稳定, 鲁棒性强.

1 方法

将深度卷积神经网络模型与高清卫星影像数据相结合, 充分挖掘卫星遥感数据的价值, 识别不同年份、不同拍摄角度、不同清晰度下的卫星图片上的

变化了的建筑物像元图斑. 其端到端的检测系统如图 1 所示, 给定包含图像与相应真实标签的数据集, 训练深度学习网络, 使其能够直接从输入图像中预测出变化了的建筑物.

将深度学习与建筑物变化检测相结合的整体方案如图 2 所示, 该方案的实施分为三个阶段: 数据预处理、模型训练和预测结果. 通过对原始图像进行预处理制作数据集, 并划分为训练集和测试集, 然后将训练集图像输入到 FlowS-Unet 网络中, 进行充分训练后得到优化的网络模型. 最后将测试图像输入训练好的模型中, 根据最终输出得到预测结果.

1.1 数据预处理

本实验数据从天池大数据平台获得, 分别为: 1) 2015 年与 2017 年广东省某地的卫星遥感数据, 所有影像由多景数据拼接而成, 覆盖广东省部分地区数百平方公里的土地, 是国土资源工作中常见的

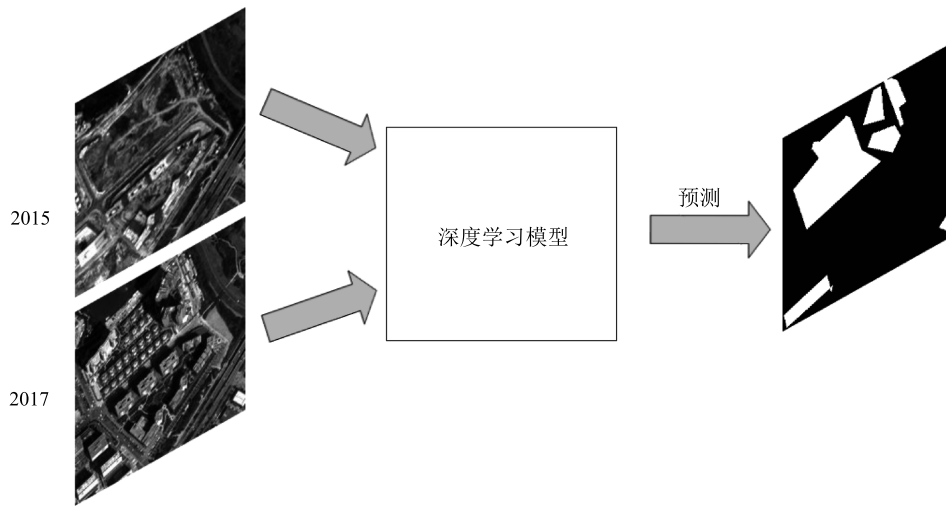


图 1 一种端到端的建筑物变化检测概览图

Fig.1 An overview of end-to-end architecture of building change detection

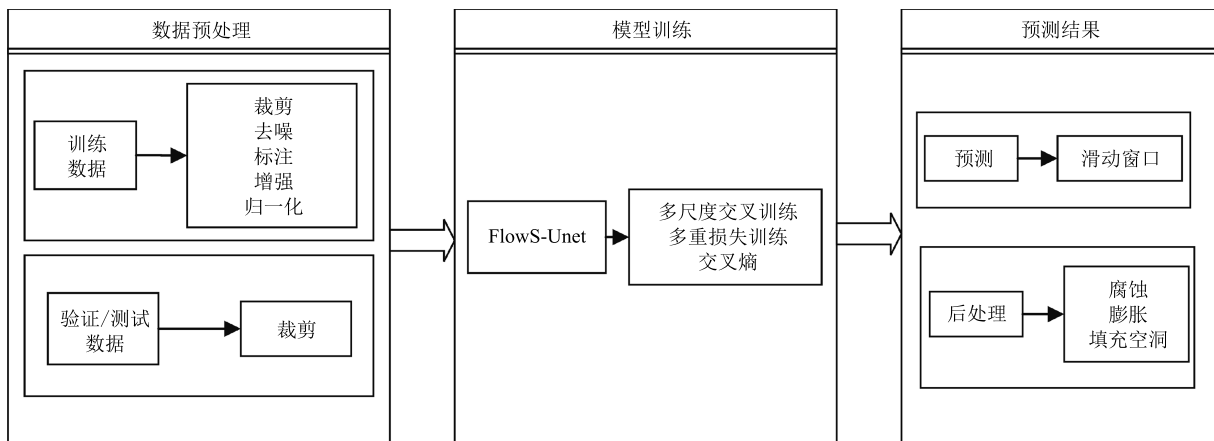


图 2 整体方案实施流程

Fig.2 The schema of proposed approach

实际场景,大小为 15 106 像素 × 15 106 像素,共四通道,每像元占 16 位,数据实际范围为 0~2 774,其中 0~450 和 1 500~2 774 占有非常少的比例,图 3 为其中部分卫星图像; 2) 包含 5% 地块的国土审批记录图; 3) 一份人工精准标注的小型数据集. 所有数据图像格式为 tiff 格式. 因未提供完整标注,所以需要自行构造训练集. 本文对该数据集做了以下几个方面的处理.



图 3 部分卫星图原图

Fig. 3 Part of the original satellite image

1) 原卫星图片上有明显的波段拼接痕迹,某些地区有明显的云、雾覆盖现象,这对识别新增建筑物的精度都有很大的干扰. 所以,本文使用 Canny 算法^[26]检测到图像拼接边缘后,将原图分为四个部分,每次提取一块作为感兴趣区域,对其他区域进行屏蔽,生成掩膜 (Mask) 图像,再对每幅掩膜图像中感兴趣区域进行 2% 的线性拉伸^[27],即将直方图累积为 2% 处的像元值作为下限 MinValue,将直方图累积为 98% 处的像元值作为上限 MaxValue,如果像元值大于 MinValue 且小于 MaxValue,则将其拉伸为 0~255; 如果像元值小于 MinValue,那么将其改为 0; 如果像元值大于 MaxValue,那么将其改为 255,以此来改善拼接区域颜色差异大的问题,改善的效果如图 4. 对于有云雾现象,参考 Cai 等^[28]提出的 DehazeNet,对其进行微调,使得参数适应本文的数据,智能学习云雾特征极大地实现去云、雾功能,其效果如图 5 所示. 对于不同卫星因不同拍摄角度导致的建筑倾斜角度不同的问题,未有可靠的预处理方式报道. 本文设计的网络初衷就是让模型具有自主学习能力,找到两个输入图像的对应关系,对不同地区不同卫星拍摄的卫星图上检测建筑物变化能具有较高的准确率,不受不可抗因素干扰的影响,能够具有很好的泛化能力,本文实验结果也证明本模型具有这种能力.

2) 使用自制标注工具对整个图像进行人工标注,绘制多边形来覆盖变化的建筑物,部分标记结果如图 6 所示. 然后对标签进行 one-hot 处理,使之成为二通道数据,即将变化建筑物用 1 表示,其他则用 0 表示,与原图一起作为模型的输入. 将整个图像分为训练集、验证集以及测试集三部分,大小分别为 15 106 像素 × 11 106 像素、15 106 像素 × 1 000 像

素和 15 106 像素 × 3 000 像素. 对训练集采取多尺度图像块与滑动窗口相结合,利用区域块重叠技术覆盖到整个图像,将图像切割为 224 像素 × 224 像素、256 像素 × 256 像素、288 像素 × 288 像素和 320 像素 × 320 像素大小的图像块,再经过旋转、翻转和缩放等进行数据增强,来确保模型的识别精度和泛化能力. 验证集和测试集则未进行数据增强.

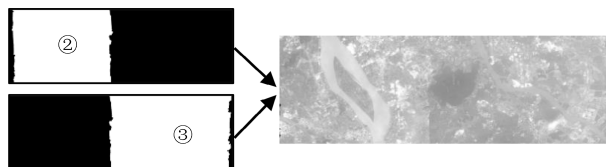


图 4 减少拼接区域色差效果图

Fig. 4 The result after the color difference of stitching blocks was suppressed

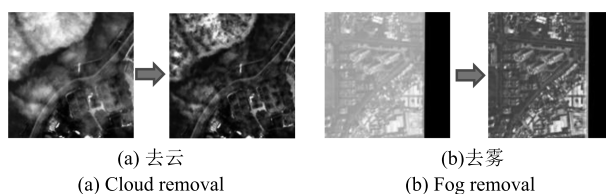


图 5 去云、雾效果图

Fig. 5 Results of cloud and fog removal



图 6 人工标注示意图

Fig. 6 The demonstration of manual label annotation

3) 因数据不同区域分布有很大的不一致性,遥感图像的平均亮度、像素值的分布范围有比较大的差异,所以对所有数据进行归一化操作,尽可能使数据有相似的分布,减少这些差异,加快训练模型时收敛的速度,并提高模型检测的精度. 本文选择使用 z-score 标准化方法,即通过计算数据的均值和方差来进行数据的归一化. 经过处理的输入图像数据将符合标准正态分布,即均值为 0、方差为 1. 其计算式为

$$x' = \frac{x - \mu}{\delta} \quad (1)$$

其中, μ 为所有样本数据的均值, δ 为所有样本数据的方差.

1.2 FlowS-Unet 网络模型结构

所构建的网络命名为 FlowS-Unet, 它不仅能提

取建筑物特征, 还能检测出图像序列中建筑物对应像素在时间域上的变化, 从而检测出建筑物变化, 模型总体结构如图 7 所示.

该网络呈对称结构, 输入图像与输出图像大小相等, 卷积层的数量有 17 个, 共进行 3 次下采样, 3 次上采样操作, 无全连接层, 每一层产生的特征都含有加强的语义信息.

网络左半部分是卷积神经网络中的一种典型结构, 通过逐渐缩减输入数据的空间维度, 来提取高层特征. 核心是 4 组卷积操作, 同一组内采用 2 次连续的 3×3 卷积, 卷积核数目相同. 从较浅组至较深组, 卷积核的数目从 32 依次成倍增多扩至 256. 每两组之间做池化操作进行空间降维, 过滤掉一些不重要的高频信息, 特征图的空间尺寸会相应缩小. 卷积后为了减少因网络加深在训练阶段出现的梯度弥散现象, 采用 ReLU^[29] 函数作为所有激活函数. 同时在下游路径添加了 BN (Batchnorm) 层, 让每个 Batch 对网络中的每一层特征都做归一化, 使得每层分布相对稳定, 加快模型收敛速度并提升模型的容纳能力^[30].

右半部分自底向上由反卷积层^[31]组成, 以图 7 虚线框中的层为例, 输入包括三部分: 上一层进行反卷积操作得到的两种深层抽象特征, 以及相应的网络左半部分生成的含有浅层局域特征的特征. 将这三个输入特征通过通道连接的方式进行融合, 从

而逐步恢复目标的细节和相应的空间维度. 并在合并路径上添加了 Dropout 层, 随机让 50% 的隐藏节点权重不工作, 以此提高网络泛化能力, 一定程度上防止过拟合^[32]. 融合后为消除上采样的混淆效应, 采用 3×3 的卷积核进行两次卷积. 为了每层都能进行独立预测, 使用 2 个 1×1 的卷积核对此时的特征向量映射到期望的类别数上. 重复上述操作三次后, 最后一层输出图像可恢复到与输入图像相同的尺寸, 并对每个像素进行分类预测. 对于网络模型优化策略上, 我们还做了以下一些细节方面的改进:

1) 多尺度交叉训练

全卷积网络的优势之一就是可接受任意尺寸的输入图像. 实验发现, 因需对数据进行裁剪, 若用单一尺寸的图像, 对处于裁剪边缘的建筑物变化很难实现正确预测, 尤其是当同一建筑物被裁剪到两张不同图片上时, 这严重影响边缘部分的预测精度. 因此, 本文将原图通过滑动窗口切片成 224 像素 \times 224 像素、 256 像素 \times 256 像素、 288 像素 \times 288 像素和 320 像素 \times 320 像素四种不同尺寸, 将不同尺寸的图像组成不同的 Batch, 放入 FlowS-Unet 中进行交叉训练, 这样可在节省内存的同时充分利用数据集, 以便对不同尺度的建筑物进行完整分割, 提高分割精度, 得到更好的识别结果.

2) 多重损失

底层特征含有丰富的位置信息, 对检测变化的

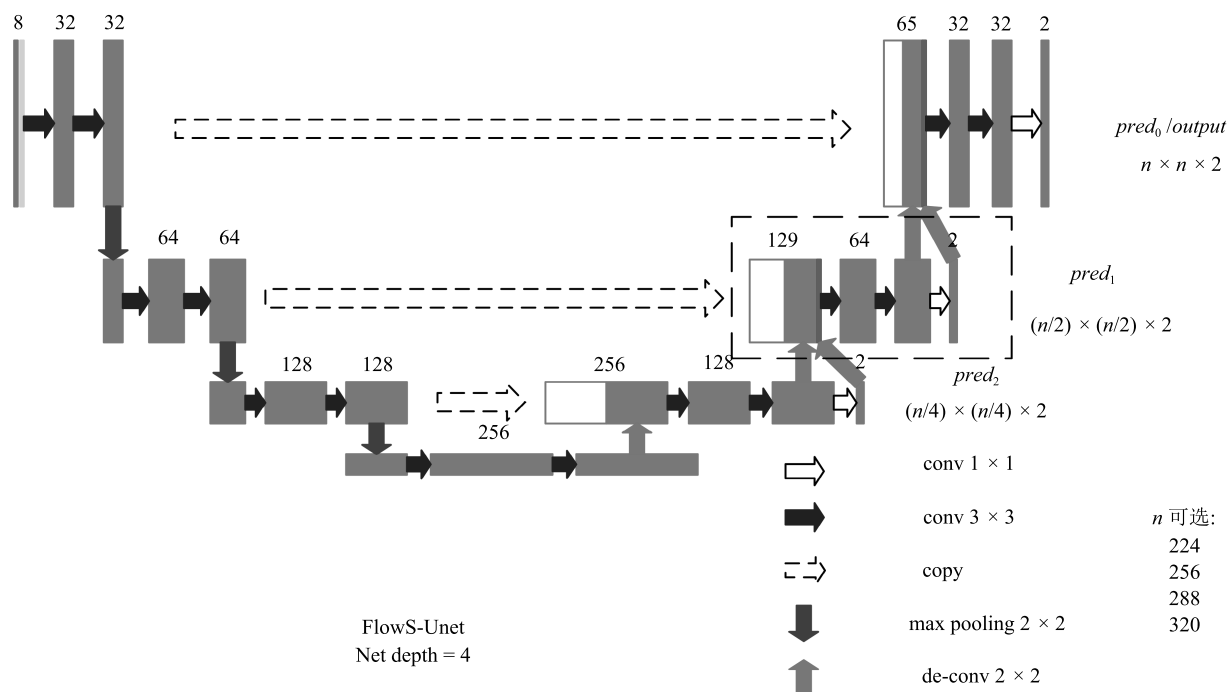


图 7 FlowS-Unet 网络结构

Fig. 7 The network structure of FlowS-Unet

建筑物尤其是小建筑是很有帮助的. 因此, FlowS-Unet 将反卷积特征与浅层特征串接后, 根据该融合特征对每一层进行了独立预测, 得到 $pred_0, pred_1, \dots, pred_n$ 多个预测结果再求取这多个预测的加权和^[33]. 由于除最后一层外, 其他层输出特征图都比实际的标签图像要小, 因此计算损失值时需将标签图像缩小至与该层输出特征图大小相同. 将这 n 个权重均设置为 1, 学习率初始为 0.001, Adam 作为优化器^[34], 总的损失函数公式如下:

$$Loss = W_0L(y, pred_0) + W_1L(y, pred_1) + \dots + W_nL(y, pred_n) \quad (2)$$

其中, $W_i, i = 1, \dots, n$, 为损失权重 (本文均设置为常数 1), n 为网络深度.

3) 交叉熵

在神经网络反向传播中, 需要一个损失函数来表示真实值与网络预测值的误差, 才能在迭代过程中优化网络权重. 本文单层损失函数使用交叉熵代价函数, 如式 (3), 它具有非负性, 且当真实输出与期望输出接近时代价函数更接近于 0. 当误差大的时候, 权重更新快, 当误差小的时候, 权重的更新可放慢. 前向传播至计算出总损失值后, 才进行反向传播.

$$C = -\frac{1}{N} \sum_n y \ln a + (1 - y) \ln(1 - a) \quad (3)$$

其中, y 为期望的输出, a 为神经元实际输出.

1.3 预测结果的后处理

上述网络的预测结果中仍然存在孤立点、空洞等噪声, 为使预测结果更精确, 必须采用恰当的方法去除这些干扰. 本文采用膨胀、腐蚀和孔洞填充等方法来细化预测结果 (如图 8).

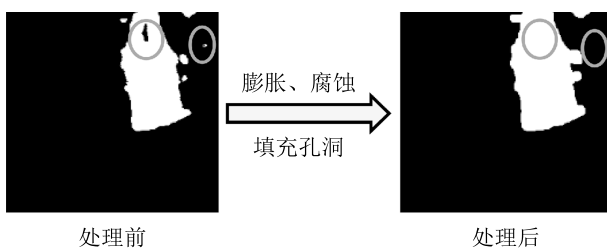


图 8 后处理前后对比图

Fig. 8 A comparison between the results before and after post-processing

1) 膨胀

膨胀是将与物体接触的所有背景点合并到物体中, 使边界向外扩张的过程. 由于图像本身就带有噪声, 预测出的边界很不平滑, 使用膨胀操作可在不

明显改变建筑物所占面积的情况下, 消除单个块边缘预测很弱的现象, 平滑其边界, 测出的边缘更清晰化. 本文使用 OpenCV 中的 Dilation 函数, 其中操作内核参数设置为简单的 3×3 矩阵, 对图像连续膨胀 5 次.

2) 腐蚀

腐蚀是一种消除边界点, 使边界向内收缩的过程, 可以消除小且无意义的孤立区域. 通过腐蚀操作去除小于给定阈值面积的连通区域, 把检测出的过小建筑物移除, 在纤细点处分离开不同建筑物. 通过计算, 训练集标签中最小建筑物标记面积为 200, 于是把阈值面积定为 200, 即在预测结果上将面积小于 200 的建筑物标记腐蚀, 达到理想效果.

3) 孔洞填充

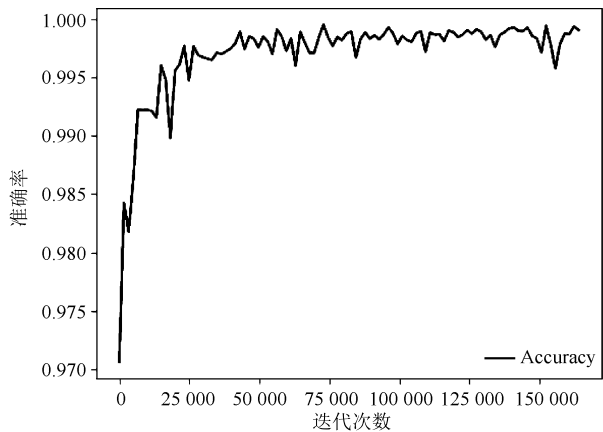
我们查看预测结果发现, 在极少大建筑物内, 出现封闭孔洞现象. 本文基于漫水填充算法 (Flood-fill), 对二值图像中的封闭孔洞进行填充, 完善所得预测结果. 该算法主要思路为将原图像向外扩展两个像素, 填充值为背景色 (黑色); 使用 OpenCV 中的 floodFill 函数, 对图的背景全部用前景色 (白色) 填充, 即得到孔洞图, 此时孔洞用黑色表示; 将填充好的图像裁剪为原图像大小后取反并与原图像相加即得到最后填充好的图像. 其中 floodFill 函数的作用是用我们指定的颜色从种子点开始填充一个连通域. 该函数第 1 个参数为输入单通道图像, 第 2 个参数为填充算法起始点设置为坐标 (0,0), 第 3 个参数为像素点被染色的值, 即在重绘区域像素的新值设置为 255, 其他参数设为默认值.

2 实验结果与分析

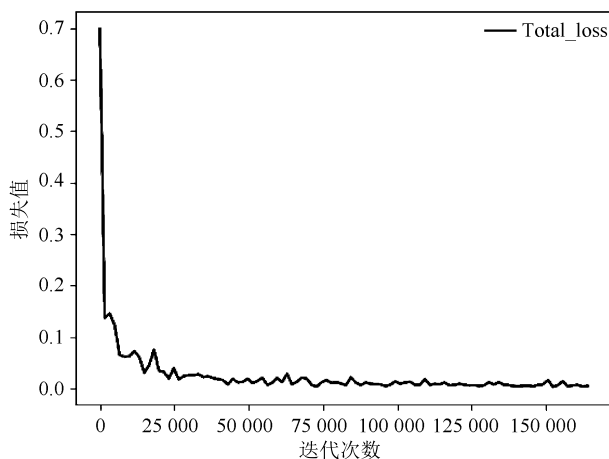
本文实验的硬件环境为 CPU Core (TM) 3.40 GHz, GPU NVIDIA GeForce GTX 980Ti, 内存 16 GB, 显存 6 GB. 软件环境为 Python 3.5, TensorFlow 1.4.0. 实验数据已在第 2.1 节详细描述, 所得的数据集以接近于 10:1:2 的比例分为训练集、验证集与测试集, 三个集合之间无相互重叠区域. 对于检测建筑物变化的需求, 一个简单的思路就是将两年的数据分别使用分类网络进行特征学习, 预测出所有的建筑物, 再将预测结果相减获得变化建筑标记. 但由于实验卫星图数据是在不同时期、不同角度下进行拍摄的, 存在一定的位置偏移和多景数据的拼接痕迹, 使得上述思路难以奏效, 而且选用该传统匹配算法时, 标注标签时需把所有建筑物均标记出来, 非常费时费力, 预测出的结果还含有许多噪声, 后期很难去除. 而使用 FlowS-Unet 模型, 则将训练集数据处理好后, 将两期卫星图合并为 8 通道数据作为模型输入, 对标签进行 one-hot 编码处理, 通过 Adam 算法不断进行迭代优化, 找到全局最优

点, 得到最佳效果. 随着迭代次数增多, 准确率不断上升, 损失值不断下降并趋于平稳后, 网络收敛, 训练结束.

图 9(a) 为训练过程中验证集上的准确率变化曲线, 图 9(b) 为损失值变化曲线图, 由图 9 可知网络在迭代 160 000 次之后趋于收敛. 在预测时, 对测试数据进行滑动窗口预测, 窗口大小为 256 像素 \times 256 像素, 步长为 150, 以尽可能地覆盖建筑物边缘部分. 本文采用式 (4) 表示的 F1 分数指标作为评价标准, 又称平衡 F 分数 (Balanced F score), 它被定义为精确率和召回率的调和平均数, 即将标签中非 0 像素点区域作为正样本, 0 像素点区域作为负样本, 对模型识别分割的精度进行评估, 每个像素点的分类结果都会影响得分. 该值得分越高, 表示分割效果越好. 因此, 本文使用膨胀、腐蚀以及孔洞填充三个方法对网络得到的预测结果进行后处理, 处理结果如图 8 所示.



(a) 准确率
(a) accuracy



(b) 损失
(b) loss

图 9 准确率与损失值曲线

Fig. 9 Curves of accuracy and loss

$$F1 = 2 \times \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

其中, *precision* 为准确率, *recall* 为召回率.

本文以人工标注结果为参考标准, 选取 15 106 像素 \times 3 000 像素的卫星图片作为测试集, 与 FCN、U-Net 等网络作对比, FCN 和 U-Net 输入数据与本文相同. 表 1 显示三种方法采取后处理操作前后时计算得出的 F1 分数与模型进行分割所消耗的时间. 由表 1 可发现, 使用 FlowS-Unet 与其他自动检测方法相比, 检测效果更好, 检测所需时间与 FCN 和 U-Net 相当, 而相对于人工标注而言, 则明显省时省力, 能达到自动检测建筑变化的目的. 表 2 是使用 224, 256, 288, 320 这四种尺度训练与多尺度交叉训练的 F1 分数比较, 并使用不同尺度窗口进行预测, 其中预测时的多尺度是指 4 种大小窗口预测所得概率值求平均后进行分类, 表 2 数据表明, 多尺度交叉训练使预测准确度有明显提升, 同时对输入图像大小变化有较好的宽容度, 多尺度窗口预测也有助于提高检测正确率. 表 3 为 2017 年阿里云天池举办的“广东政务数据创新大赛—智能算法赛”的线上评测初赛 (9 月 11 日~11 月 3 日)、复赛 (11 月 6 日~11 月 12 日) 以及决赛 (11 月 21 日~22 日) 中使用不同区域的数据集上综合前 5 名所得的 F1 分数对比结果, 其中 FlowS-Unet 取得了综合第 2 名. 由于自行标注的标签与平台使用的 Ground Truth 存在一定的差别, 因此所得的 F1 分数存在少量不一致之处. 赛中本队对 FlowS-Unet 预测的结果进行了后处理操作.

表 1 FlowS-Unet 与现有方法的性能比较

Table 1 The performance comparison of FlowS-Unet

序号	方法	F1 分数 (后处理前/后)	时间 (s)
1	FlowS-Unet	0.933/0.943	62
2	FCN	0.858/0.873	50
3	U-Net	0.898/0.913	59
4	人工标注	1.000	18 000

为了更直观地展示分割效果, 图 10 显示出使用不同模型对卫星图进行建筑物变化检测后, 进行后处理操作前后的相应结果比较, 与参考标准比较后可发现, FlowS-Unet 所得结果在精确度、边缘明晰度上要明显优于基于 FCN 和 U-Net 等方法, 这些方法特征表达能力较弱, 分割出的结果噪点较多, 且建筑物分割不太完整, 边缘的错误比较显著. 图 10 也

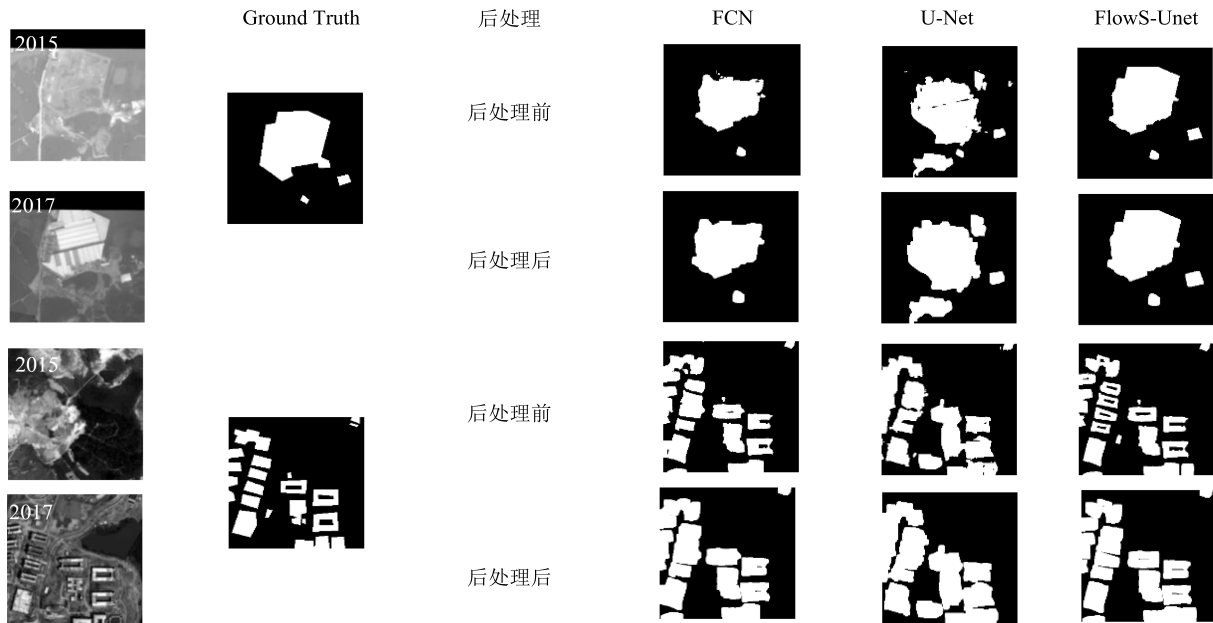


图 10 FlowS-Unet 与现有方法的定性比较

Fig. 10 A quality comparison of FlowS-Unet and previous methods

表 2 多尺度与单尺度训练及预测的 F1 分数比较

Table 2 The F1 score comparison between multi-scale cross and single-scale training and testing

训练尺度 (像素)	预测尺度 (像素)	F1 分数 (后处理前/后)
224	224	0.903/0.923
256	256	0.909/0.928
288	288	0.913/0.931
320	320	0.911/0.932
多尺度	224	0.933/0.939
	256	0.938/0.943
	288	0.939/0.945
	320	0.939/0.944
	多尺度平均	0.942/0.946

表 3 FlowS-Unet 与其他队伍的 F1 分数比较

Table 3 The F1 score comparison of FlowS-Unet and other teams

名次	初赛	复赛	决赛
第 1 名	0.890	0.914	0.861
第 2 名 (FlowS-Unet)	0.903	0.877	0.840
第 3 名	0.867	0.898	0.800
第 4 名	0.706	0.870	0.842
第 5 名	0.879	0.936	0.823

表明后处理中的空洞填充算法对于一些真实状态就

是带空洞的建筑物是会降低分割性能的,降低 F1 分数,如图 10 第 2 幅图像,本是新造地基形成的真实孔洞,但使用空洞填充算法后该孔洞就被填充完整,而未进行后处理的结果反而更接近于真实标签.但建筑物变化检测的主要目的是自动检测出建筑物的新建、扩建和改建等变化情况,至于建筑物本身不带空洞,对检测结果并无太大影响.

3 结束语

针对目前人为探查土地资源利用情况的任务具有人力成本高、劳动强度大、办事效率很低下等问题,本文提出了一种基于卷积神经网络的智能建筑物变化检测方法,将深度卷积神经网络模型与高清卫星影像数据相结合,充分挖掘卫星遥感数据的价值,助力解决城市治理的痛点难点,帮助政府实现信息化向智能化的转变,具有较高的社会价值与经济价值.一方面通过对卫星图像进行裁剪、去噪、归一化、数据增强等方式,处理好的数据通过多尺度特征融合与滑动窗口相结合的方式输入网络中.另一方面,在全卷积神经网络的基础上提出 FlowS-Unet 模型,很好地将低层高分辨率信息与高层语义信息融合到一起,并在每个融合后的特征图上单独进行预测,来减小损失值,增强网络的表达能力.最后对网络预测的结果进行膨胀、腐蚀以及填充孔洞等一系列的后处理,使得分割效果更加完美.通过与传统的人工标注结果对比发现,该方法能够快速地识别出不同年份、不同拍摄角度、不同清晰度下的卫星图片中的建筑物方面的差异性,并完整地将目标自

动分割出来, 具有高效性、可实施性. 并且此方法还可延伸到相类似的图像匹配方面, 如应用到医疗图片的分析与识别、纸币等印刷品的缺陷检测等, 只要稍加修改就可以对其他领域进行研究和扩展, 具有非常强的可塑性. 因硬件设备计算能力有限, 边缘分割还不够精细, 训练模型的时间性能还有提升. 如何在保证精确检测出建筑物变化的情况下提高其时间性能是下一步主要研究的内容.

References

- 1 Beumier C, Idrissa M. Building change detection from uniform regions. In: Proceedings of the 17th Iberoamerican Congress Pattern Recognition, Image Analysis, Computer Vision, and Applications. Buenos Aires, Argentina: Springer, 2012. 648–655
- 2 Turker M, Sumer E. Building-based damage detection due to earthquake using the watershed segmentation of the post-event aerial images. *International Journal of Remote Sensing*, 2008, **29**(11): 3073–3089
- 3 Huang X, Zhang L P, Zhu T T. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, **7**(1): 105–115
- 4 Zhou Ze-Ming, Meng Yong, Huang Si-Xun, Hu Bao-Peng. Building segmentation of spaceborne SAR images based on energy minimization. *Acta Automatica Sinica*, 2016, **42**(2): 279–289
(周则明, 孟勇, 黄思训, 胡宝鹏. 基于能量最小化的星载 SAR 图像建筑物分割方法. *自动化学报*, 2016, **42**(2): 279–289)
- 5 Li Wei-Ming, Wu Yi-Hong, Hu Zhan-Yi. Urban change detection under large view and illumination variations. *Acta Automatica Sinica*, 2009, **35**(5): 449–461
(李伟明, 吴毅红, 胡占义. 视角和光照显著变化时的变化检测方法研究. *自动化学报*, 2009, **35**(5): 449–461)
- 6 Tian Hao, Yang Jian, Wang Yan-Ming, Li Guo-Hui. Towards automatic building extraction: Variational level set model using prior shape knowledge. *Acta Automatica Sinica*, 2010, **36**(11): 1502–1511
(田昊, 杨剑, 汪彦明, 李国辉. 基于先验形状约束水平集模型的建筑物提取方法. *自动化学报*, 2010, **36**(11): 1502–1511)
- 7 Liu B, Tang K, Liang J. A bottom-up/top-down hybrid algorithm for model-based building detection in single very high resolution SAR image. *IEEE Geoscience and Remote Sensing Letters*, 2017, **14**(6): 926–930
- 8 Lukashovich P, Zalessky B, Belotserkovsky A. Building detection on aerial and space images. In: Proceedings of the 2017 International Conference on Information and Digital Technologies (IDT). Zilina, Slovakia: IEEE, 2017. 246–251
- 9 Shi Wen-Zao, Mao Zheng-Yuan. The research on building change detection from high resolution remotely sensed imagery based on graph-cut segmentation. *Journal of Geo-Information Science*, 2016, **18**(3): 423–432
(施文灶, 毛政元. 基于图分割的高分辨率遥感影像建筑物变化检测研究. *地球信息科学学报*, 2016, **18**(3): 423–432)
- 10 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Nevada, USA: Curran Associates Inc., 2012. 1097–1105
- 11 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, California, USA, 2015. 1–14
- 12 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. 3431–3440
- 13 Yuan J Y. Automatic building extraction in aerial scenes using convolutional networks. arXiv: 1602.06564, 2016.
- 14 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015. 234–241
- 15 Hariharan B, Arbeláez B, Girshick R, Malik J. Hypercolumns for object segmentation and fine-grained localization. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. 447–456
- 16 Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V, et al. FlowNet: Learning optical flow with convolutional networks. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 2758–2766
- 17 Chen Wen-Kang. Remote sensing image detection of rural buildings based on deep learning algorithm. *Surveying and Mapping*, 2016, **39**(5): 227–230
(陈文康. 基于深度学习的农村建筑物遥感影像检测. *测绘*, 2016, **39**(5): 227–230)
- 18 Silberman N, Sontag D, Fergus R. Instance segmentation of indoor scenes using a coverage loss. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 616–631
- 19 Farabet C, Couprie C, Najman L, LeCun Y. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(8): 1915–1929
- 20 Zhang A, Liu X M, Gros A, Tiedecke T. Building detection from satellite images on a global scale. arXiv: 1707.08952, 2017.
- 21 Ghaffarian S, Ghaffarian S. Automatic building detection based on supervised classification using high resolution Google earth images. In: Proceedings of the 2014 ISPRS Technical Commission III Symposium. Zurich, Switzerland: ISPRS, 2014. 101–106
- 22 Shu Z, Hu X Y, Sun J. Center-point-guided proposal generation for detection of small and dense buildings in aerial imagery. *IEEE Geoscience and Remote Sensing Letters*, 2018, **15**(7): 1100–1104

- 23 Yang H L, Lunga D, Yuan J Y. Toward country scale building detection with convolutional neural network using aerial images. In: Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth, TX, USA: IEEE, 2017. 870–873
- 24 Sun L, Tang Y Q, Zhang L P. Rural building detection in high-resolution imagery based on a two-stage CNN model. *IEEE Geoscience and Remote Sensing Letters*, 2017, **14**(11): 1998–2002
- 25 Vakalopoulou M, Bus N, Karantzas K, Paragios N. Integrating edge/boundary priors with classification scores for building detection in very high resolution data. In: Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth, TX, USA: IEEE, 2017. 3309–3312
- 26 Canny J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, **PAMI-8**(6): 679–698
- 27 Wei Ya-Xing, Wang Li-Wen. Analysis of enhancement methods about satellite images. *Geomatics and Spatial Information Technology*, 2006, **29**(2): 4–7
(卫亚星, 王莉雯. 遥感图像增强方法分析. 测绘与空间地理信息, 2006, **29**(2): 4–7)
- 28 Cai B L, Xu X M, Jia K, Qing C M, Tao D C. DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 2016, **25**(11): 5187–5198
- 29 Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. *Journal of Machine Learning Research*, 2011, **15**: 315–323
- 30 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR, 2015. 448–456
- 31 Zeiler M D, Krishnan D, Taylor G W, Fergus R. Deconvolutional networks. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 2528–2535
- 32 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014, **15**(1): 1929–1958
- 33 Lin T Y, Dollár P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 936–944
- 34 Kingma D P, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA: 2015. 1–15



顾炼 浙江工商大学计算机与信息工程学院硕士研究生. 主要研究方向为图像处理, 模式识别.

E-mail: guliencv@163.com

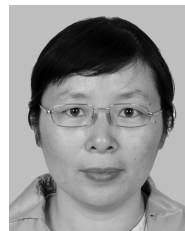
(GU Lian Master student at the School of Computer and Information Engineering, Zhejiang Gongshang University. Her research interest covers image processing and pattern recognition.)



许诗起 浙江工商大学计算机与信息工程学院硕士研究生. 主要研究方向为数据挖掘, 深度学习.

E-mail: xushiqtc@163.com

(XU Shi-Qi Master student at the School of Computer and Information Engineering, Zhejiang Gongshang University. His research interest covers data mining and deep learning.)



竺乐庆 浙江工商大学计算机与信息工程学院副教授. 主要研究方向为图像处理, 模式识别, 视频处理. 本文通信作者.

E-mail: zhuleqing@zjgsu.edu.cn

(ZHU Le-Qing Associate professor at the School of Computer and Information Engineering, Zhejiang Gongshang University. Her research interest covers

image processing, pattern recognition, and video processing. Corresponding author of this paper.)