



# 基于 Mask R-CNN 的遥感影像土地分割与轮廓提取

## 摘要

传统的模式识别难以对土地遥感影像一次性精确统计分析.在精确分割出土地种类的前提下,本文提出了一种基于骨干网络为 ResNet-101-RPN 的 Mask R-CNN 的遥感影像土地分割与轮廓提取方法.该方法包括以下步骤:数据获取、图像去雾、遥感影像土地统计分析、土地分割和轮廓提取.在一个具有挑战性的卫星地图瓦片数据集上对所提出的方法进行训练和测试.实验结果表明,该方法以 0.907 的均值平均精度 (mAP) 和 31.33 像素的均值平均距离误差 (mADE) 获得了令人满意的多种土地分割和轮廓提取结果.

## 关键词

土地分类;轮廓提取;地图瓦片;Mask R-CNN

中图分类号 TP391

文献标志码 A

收稿日期 2020-09-10

资助项目 南方电网公司科技项目(030700KK52190003)

## 作者简介

黎江,男,硕士生,主要研究方向为图像处理技术.251057729@qq.com

许明慧(通信作者),女,硕士生,研究方向为鸟害风险预测技术.1842223946@qq.com

1 广东工业大学 自动化学院,广州,510006

2 广东电网有限责任公司 江门供电局,江门,529000

## 0 引言

传统的城镇空间规划主要依靠对多波段遥感影像分析结果,以多光谱的 Landsat-8 遥感影像为代表<sup>[1]</sup>.高分辨率的遥感影像具有真实、实时以及便捷等特点,精确的图像分割是统计分析的前提.图像边缘检测利用图像的边缘信息实现图像分割是图像处理研究的一个基础方向,相关算法有 LoG 算子<sup>[2]</sup>、Sobel 算子<sup>[3]</sup>、Canny 算子<sup>[4]</sup>、Roberts 算子<sup>[5]</sup>等.加拿大的 Hay 等<sup>[6]</sup>提出了多尺度分割方法对卫星影像进行分割;彭志平等<sup>[7]</sup>针对我国常用地图中的棕、蓝、绿、黑 4 种颜色在彩色空间的分布规律提出了快速阈值分割算法.为了提高图像分割效率,比利时的 Vancoillie 等<sup>[8]</sup>在处理图像分割时使用了遗传算法以及神经网络.

近年来,具有较强特征学习能力的深度卷积神经网络(CNNs)广泛应用在图像分割领域<sup>[9-10]</sup>.CNNs 通过大量标注图像数据的训练,利用学习到的这些携带了丰富的语义和空间信息的特征去进行图像分割<sup>[11-12]</sup>.此外,DeepMask 和 SharpMask<sup>[13]</sup>直接从图像像素生成分割对象,然后将其分类为不同的类别.在 Faster R-CNN<sup>[14]</sup>算法的基础上,结合经典语义分割算法 FCN<sup>[15]</sup>,He 等<sup>[16]</sup>提出了用于实例分割与关键点检测的 Mask R-CNN,该网络框架由 4 部分组成,分别是卷积神经网络 CNN、区域建议网络、ROIAlign 与输出模块,使像素分类和提取的准确率得到极大提高,研究结果表明基于 Mask R-CNN 的遥感影像土地分割和轮廓提取是可行的.

为了在遥感影像上实现一次性对土地的高精度分割和轮廓提取,本文提出一种基于骨干网络为 ResNet-101-RPN 的 Mask R-CNN 的遥感影像土地分割与轮廓提取方法.如图 1 所示,整个方法包括几个步骤:首先选用 17 级谷歌地图(Google Maps)瓦片作为训练和测试数据集;其次根据谷歌地图瓦片的特点利用暗通道去雾算法进行去雾处理获得无雾图像;然后由土地统计分析确定分割对象;最后进行土地分割和轮廓提取.在对遥感影像进行一次精确统计分析土地时,本文所提出的方法有效地解决了在有雾的地图瓦片图像上实现精确的土地分割与轮廓提取问题.

## 1 数据材料

### 1.1 数据采集

本文的模型训练和测试是在地图数据上进行的,使用目前应用

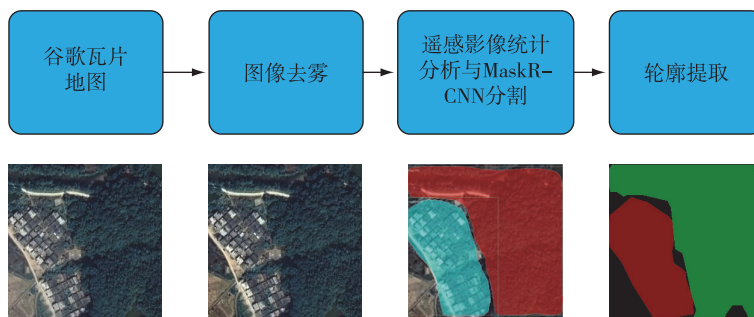


图1 基于 Mask R-CNN 的遥感影像土地分割与轮廓提取流程

Fig.1 Flow chart of remote sensing image segmentation and contour extraction based on Mask R-CNN

最为广泛的 Google 地图瓦片方案. Google 地图分为混合地图、普通地图以及卫星地图. 瓦片数据使用金字塔方式组织, Google 地图向用户提供 0~22 缩放等级, 缩放等级越高, 所包含地理范围越小, 但是包含的地理要素却越精细. 每张地图瓦片保存格式为 PNG 和 JPG, 像素大小为 256×256.

选取训练数据不但要考虑图像的清晰程度, 同时也要考虑各类数据的平衡性, 这样训练出来的模型才具有更好的泛化能力, 因此要遵循以下原则: 第一, 选取的地图瓦片要具有适应性和代表性; 第二, 土地分类数据要平衡; 第三, 一张地图瓦片中要尽量包含多种土地类别, 如选取的地图瓦片要包含森林、城镇、耕地等土地要素. 本文采集的遥感影像是缩放等级为 17 级的卫星地图瓦片, 空间分辨率 2.28 m, 即一个像素点表示地面 2.28 m×2.28 m 的面积, 图片格式为 JPG. 针对研究区域采集 1 000 张地图瓦片作为训练集, 100 张作为测试集, 使用标注工具 Labelme 对数据集样本进行分割掩码标记, 完成图像中土地的标注.

## 1.2 图像去雾

由于受到大气中雾霾细小颗粒、水蒸气等杂质影响, 卫星从高空直接拍摄得到的地图瓦片图像会出现模糊、不清晰的现象, 如果直接训练模型和测试不容易取得理想的结果, 因此为了避免这种不良影响, 首先要对采集的图像进行预处理. 结合 Google 地图瓦片的特点, 本文采用文献[17]算法对图像进行去雾处理.

### 1.2.1 暗通道先验

暗通道假设图像中非天空彩色区域每个像素只有一个很低的颜色通道值, 对于任意输入的一个地图瓦片图像, 暗通道的数学模型:

$$J^{\text{dark}}(x) = \min_{y \in \Omega(x)} \left( \min_{c \in (r, g, b)} J^c(y) \right), \quad (1)$$

在式(1)中,  $J^c$  为非天空彩色图像的 3 个不同颜色分量,  $\Omega(x)$  是一个以像素  $x$  为中心的窗口. 通过对大量非天空无雾彩色图像统计表明  $J^{\text{dark}} \rightarrow 0$ , 这就是暗通道先验定理.

### 1.2.2 大气散射模型

大气散射模型:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (2)$$

在图像处理领域被广泛应用. 其中,  $I(x)$  为待去雾的地图瓦片图像,  $J(x)$  为无雾图像,  $A$  是全球大气光照强度,  $t(x)$  为透射率, 若通过有雾的瓦片图像  $I(x)$  求出无雾的瓦片图像  $J(x)$ , 需求解  $t(x)$  和  $A$ .

将式(2)进行整理得:

$$\frac{I^c(x)}{A^c} = t(x) \frac{J^c(x)}{A^c} + 1 - t(x), \quad (3)$$

$C$  为  $r, g, b$  三通道. 由于  $A$  值可以利用暗通道从有雾地图瓦片图像获取, 假设每一个窗口透射率为  $\hat{t}(x)$ , 对式(3)作两次最小值运算:

$$\min_{y \in \Omega(x)} \left( \min_c \frac{I^c(y)}{A^c} \right) = \hat{t}(x) \min_{y \in \Omega(x)} \left( \min_c \frac{J^c(y)}{A^c} \right) + (1 - t(x)), \quad (4)$$

其中,  $J^c(x)$  是要求解的无雾瓦片图像. 所以根据暗通道先验定理得:

$$\min_{y \in \Omega(x)} \left( \min_c \frac{J^c(y)}{A^c} \right) = 0. \quad (5)$$

式(5)代入式(4)得:

$$\hat{t}(x) = 1 - w \min_{y \in \Omega(x)} \left( \min_c \frac{I^c(y)}{A^c} \right), \quad (6)$$

其中,  $w$  为  $[0, 1]$  之间的一个值, 根据 Google 地图的特点, 本文取  $w = 0.9$ .

最后将全球大气光照强度值  $A$  和  $\hat{t}(x)$  代入式(2)可求解无雾瓦片图像  $J(x)$ :

$$J(x) = \frac{I(x) - A}{t(x)} + A. \quad (7)$$

图像去雾的性能如图 2 所示.将去雾图像与原图像进行比较,可以看出居民区、森林、耕地区域明显改善,去雾后的图像清晰度得到了提高.

## 2 遥感影像土地统计分析

如果将土地详细分为 30 多类,则可以分为水田、城市、公路用地、森林、村庄、茶园、裸地、湖泊、建制镇、草地、建筑用地和沼泽地等.如果依据 30 多类对象利用 Labelme 软件标注地图瓦片,不但使标注任务艰巨,而且训练得到的模型分割和轮廓提取精度将会下降,因此要结合实际情况重新将对象细分和粗分.

### 2.1 土地对象细分和粗分

根据实际情况和实例分割使用经验,将土地对象细分为 9 类:建筑用地、道路用地、裸地、绿化、草地、森林、耕地、水域和其他,并把细分的 9 类粗分为建设用地、生态用地和耕地 3 类,其中建筑用地、道路用地、裸地以及绿化为建设用地,草地、森林、水域和其他为生态用地,剩下的是耕地.土地对象细分和粗分最终结果和说明如表 1 所示.因此图像标记对象为建设用地、生态用地、耕地.

### 2.2 土地分析

遥感影像土地统计分析需要计算研究区域的建设用地紧凑指数、生态用地斑块密度指数以及耕地破碎指数.建设用地紧凑指数(CI,其量值记为  $I_c$ )主要描述城镇建设用地的饱和以及紧凑程度,当城镇

表 1 土地分类

Table 1 Land classification

粗分类	细分类	简要说明
建设用地	建筑用地	包括高楼、民房、矮楼、场馆等建筑物
	道路用地	包括裸露的公路、桥梁和铁路
	裸地	指城市中裸露的小块和大块土地
	绿化	指城市中的绿化地区
生态用地	草地	指大片绿地
	森林	指林区的大片森林
	水域	包括湖泊和河流等
	其他	指图像边缘的留空部分
耕地	耕地	包括农田、水浇地等

的紧凑指数越大,表示其中心区域越集中和经济效益价值更大,它的数学描述为

$$I_c = \frac{2\sqrt{S_i\pi}}{P_i}, \quad (8)$$

其中,  $I_c$  为建设用地紧凑指数,  $S_i$  为城镇建设用地面积,  $P_i$  为城镇建设用地的轮廓周长.  $I_c$  在 0 和 1 之间,它的值越大,紧凑度越大,表示城镇形态越饱满.

基于像素大小为  $256 \times 256$  的地图瓦片图像计算生态用地斑块密度指数和耕地破碎指数,它们的数学描述相同,均为

$$I_{PD} = \frac{n_i}{A}. \quad (9)$$

对于生态用地斑块密度指数  $I_{PD}$ ,  $n_i$  为地图瓦片图像的生态用地块数,  $A$  为地图瓦片图像实际区域

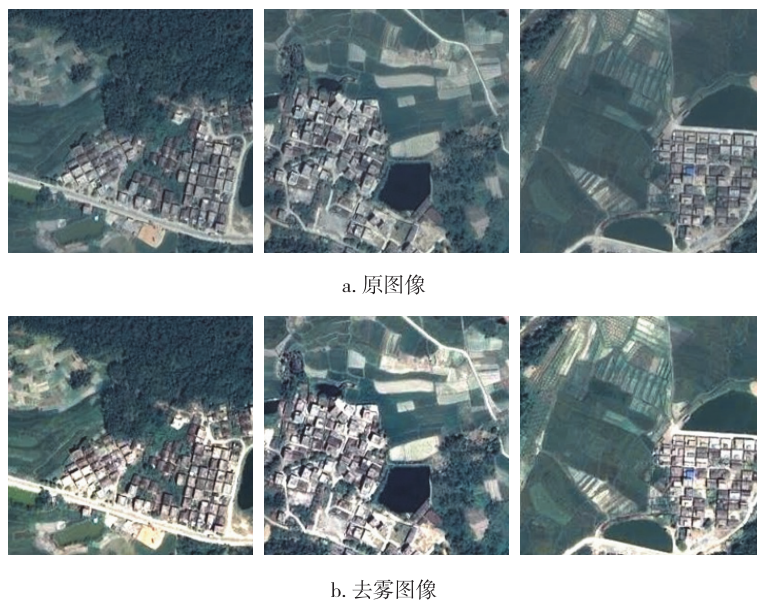


图 2 原图像与去雾图像对比

Fig. 2 Comparison between original images (a) and defogged images (b)

的生态用地面积;对于耕地破碎指数  $I_{PD}$ ,  $n_i$  为地图瓦片图像的耕地块数,  $A$  为地图瓦片图像实际区域的耕地面积.  $A$  值越小, 分别表明图像实际区域的生态用地越集中和耕地越破碎.

### 3 基于 Mask R-CNN 的土地分割与轮廓提取

本文的研究重点是将土地从地图瓦片中分割以及提取出来, 这是一个典型的实例分割任务. 基于骨干网络为 ResNet-101-RPN 的 Mask R-CNN 的土地分割与轮廓提取方法是将识别土地种类、提取不同土地像素结合起来进行的, 符合遥感影像统计分析的要求.

#### 3.1 Mask R-CNN 神经网络

Mask R-CNN 作为一种非常灵活的神经网络框架, 它在 Faster R-CNN 的基础上增加一个语义分割 FCN 分支. 基于 Mask R-CNN 的土地分割与轮廓提取框架如图 3 所示, 它包括 2 个部分: 1) 卷积骨干部分负责的是对整个地图瓦片的特征提取; 2) 头部部分执行边框识别和掩码预测. 区域建议网络 RPN 计算区域提案, 然后 RoIAlign 从每个提案中结合特征图提取特征并执行 2 个平行操作. 在全连接层 FC 进行土地的检测、分类以及边框回归. 在全卷积层 FCN 输出高精度土地分割和轮廓提取.

由于使用了高级语义信息和高分辨率特征映射, 特征金字塔网络 FPN<sup>[18]</sup> 能对土地目标进行精确定位. 为了提高模型的精度和速度, 本文使用基于 FPN 的残差网络 ResNet101<sup>[19]</sup> 作为卷积骨干.

对于每张地图瓦片图像, 首先使用 ResNet-101-RPN 提取瓦片图像的土地特征, 然后 RPN 在特征图上使用滑动窗口方法来计算边框方案. 其次, RoIAlign 通过双线性插值原理将任意大小的感兴趣空间区域映射到固定的空间分辨率. 最后, 利用 Mask R-CNN 头部预测目标类别, 细化边框定位, 同时生成土地分割掩膜.

#### 3.2 多特征提取

为了提高 Mask R-CNN 分割和土地轮廓提取的性能, 将使用更多的特征信息. 因此选择强大的 ResNet-101 作为主干. 它由 5 个阶段组成, 对应 5 个不同比例尺的特征图  $[C_1, C_2, C_3, C_4, C_5]$ . 利用这些特征映射建立 FPN 网络的特征金字塔, 分别得到新的特征图  $[P_1, P_2, P_3, P_4, P_5]$ . 使用  $P_1$  即计算  $C_1$  对应的特征映射需要花费大量的时间, 因此通过下采样  $P_5$  得到  $P_6$  来替代  $P_1$ . 特征图对应关系如下:

$$\begin{cases} P_2 = \text{conv}(\text{sum}(\text{upsample}(P_3, \text{conv}(C_2))) ), \\ P_3 = \text{conv}(\text{sum}(\text{upsample}(P_4, \text{conv}(C_3))) ), \\ P_4 = \text{conv}(\text{sum}(\text{upsample}(P_5, \text{conv}(C_4))) ), \\ P_5 = \text{conv}(\text{conv}(C_5) ), \\ P_6 = \text{downsample}(P_5) , \end{cases} \quad (10)$$

其中, conv 为卷积计算, sum 为逐元素对齐操作, upsample 和 downsample 分别为上采样和下采样.

#### 3.3 区域建议网络生成与 RoIAlign 操作

区域建议网络 RPN 收到特征图  $[P_1, P_2, P_3, P_4, P_5]$  并使用滑动窗口找到土地所在 RoI 区域. RPN 网络有两个输出, 第一个输出判断描点是背景还是土

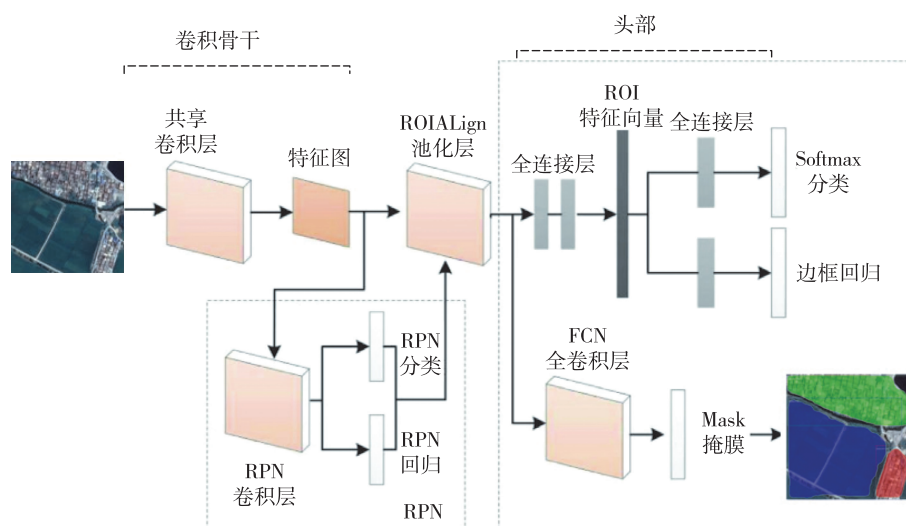


图 3 基于 Mask R-CNN 的土地分割与轮廓提取框架

Fig. 3 Framework for image segmentation and contour extraction based on Mask R-CNN

地对象,第二个输出用于校正边界框,以更好地适应土地对象.在对 RPN 网络进行处理和预测后,可以得到一系列边界框.边界框会出现相会重叠的现象,可以利用非最大值抑制(NMS)来获取土地对象分数较高的边界框并将其传递给下一级.

在头部之前,为了产生标准尺寸的输入以输入到分类器, RoIAlign 用于将描点的尺寸调整为固定大小.通过 RoIAlign 网络,将提取的特征图与输入图像进行适当地对齐,解决了区域建议与特征映射之间定义对应区域时的亚像素错位问题,使像素分割和提取更加精确,有助于获取更好的像素分割和提取结果.

### 3.4 土地实例分割与损失函数

通过 RoIAlign 网络层得到的特征被送入全连接层 FC 进行目标分类和边界框回归,同时被送入全卷积层 FCN 进行分割.分类是通过将使用所有特征的 FC 层的输出通过 Softmax 层来完成的,土地边界回归只能使用从原始区域建设中获得的特征完成,同时进行 Mask R-CNN 头部分割和提取土地区域.

对于神经网络训练,损失函数为预测值与真实值之间的差值,它在土地分割与轮廓提取模型训练中起着重要的作用.在基于 Mask R-CNN 的土地分割与轮廓提取网络中,使用联合损失函数训练边框回归、分类和掩膜预测.使用的联合损失函数如下:

$$L = L_{cls} + L_{bbox} + L_{mask}, \quad (11)$$

其中,  $L_{cls}$  为土地分类错误,  $L_{bbox}$  为土地边框回归错误,  $L_{mask}$  为掩膜分支错误.

### 3.5 土地轮廓周长以及面积获取

遥感卫星统计分析需要计算建设用地紧凑指数、生态用地斑块密度指数以及耕地破碎指数,涉及到城镇建设用地面积轮廓周长、生态用地面积和耕地面积,因此可以通过分割的土地图像来获取轮廓周长和面积.在 Mask R-CNN 实例分割中,每个分割后的图像都会被转换成一个二值矩阵,其中土地像素得分为‘1’,背景像素为‘0’,由背景像素‘0’包围的‘1’像素是土地的轮廓线,根据一个像素点代表 2.28 m×2.28 m 的区域求得土地轮廓周长和面积.

### 3.6 评估指标

采用均值平均精度(mAP,其量值记为  $\eta_{mAP}$ )评价土地分割的性能.mAP 是图像分割的一个非常流行的评估指标,根据精确率与召回率,每个类别可以绘制一条曲线,该曲线与坐标轴围成的面积就是平

均精度(AP,其量值记为  $\eta_{AP}$ ),则 AP 的平均值为 mAP,其计算公式为:

$$R = \frac{T_p}{T_p + F_N}, \quad (12)$$

$$P = \frac{T_p}{T_p + F_p}, \quad (13)$$

$$\eta_{AP} = \int_0^1 P(R) dR, \quad (14)$$

$$\eta_{mAP} = \frac{\sum \eta_{AP}}{n}, \quad (15)$$

式中,  $R$  为召回率,  $P$  为精确率,  $T_p$  为正样本被预测正确的数量,  $F_p$  为负样本被预测错误的数量,  $F_N$  为正样本被预测错误的数量,  $n$  为测试集大小.

对于土地轮廓提取,利用真值的 mADE 评估:

$$\varepsilon_{ADE} = \frac{A_{union} - A_{overlap}}{T_{contour}}, \quad (16)$$

$$\varepsilon_{mADE} = \frac{\sum_{i=1}^n \varepsilon_{ADE,i}}{n}, \quad (17)$$

式中,  $\varepsilon_{ADE}$  为真值的平均距离误差,  $\varepsilon_{mADE}$  为  $\varepsilon_{ADE}$  的均值,  $A_{union}$  是预测掩膜和真值所包围的区域,  $A_{overlap}$  为预测掩膜和标记真值之间重叠区域,  $T_{contour}$  为标记真值区域,  $n$  为测试集大小, mADE 的值越小表明轮廓提取越精确.

## 4 实验

### 4.1 实验环境

本文采用 TensorFlow<sup>[20]</sup> 结合 GPU 训练基于 Mask R-CNN 的土地分割与轮廓提取模型,本实验的硬件和软件信息如表 2 所示.

表 2 实验环境配置

Table 2 Experimental environment configuration		
系统环境	软件环境	硬件环境
Ubuntu 16.04	CUDA 10.0	Intel E5-2640×2
	OpenCV3.4.5	RAM:64 GB
	TensorFlow 框架	GTX 1070:8 GB

在整个数据集中,根据地图瓦片选取原则,一共选择 1 100 张地图瓦片.另外,根据是否进行去雾处理,有相应的原始和去雾图像数据集.在采集的数据集中,前 1 000 张地图瓦片作为训练集,剩余的 100 张作为测试集.为了满足模型训练和测试要求,利用 Labelme 软件<sup>[21]</sup> 手动标记不同种类的土地区域.此外,为了加快神经网络训练,采用基于 COCO 数据

集<sup>[22]</sup>的预训练权重.为了验证去雾图像的有效性,分别对原始图像和去雾图像数据集训练基于两个不同骨干网络的 Mask R-CNN 的分割与轮廓提取模型,两个骨干网络分别为 ResNet-101-RPN 和 ResNet-50-RPN.根据实验,设学习率为 0.001,每次迭代的批处理量为 50,该模型一共迭代了 20 000 次.

#### 4.2 实验结果与分析

原始和去雾图像的分割与轮廓提取结果图 4 所示.可以看出,在图 4a 的原图像中,骨干网络为 ResNet-50-RPN 的分割与轮廓提取结果显示无法分割识别图像中的建设用地,而在去雾图像中,骨干网络为 ResNet-101-RPN 的基本能把图像中的建设用地分割识别出来,如图 4b 所示.其主要原因是采用文献[17]算法进行去雾减轻了大气中雾霾细小颗粒、水蒸气等杂质对地图瓦片的影响,减少了对土地对象的误判,同时骨干网络 ResNet-101-RPN 有更复杂的特征提取网络.

此外,所提出骨干网络为 ResNet-101-RPN 的 Mask R-CNN 方法的一些多土地实例分割与轮廓提取结果如图 5 所示,可以看出,即使生态用地被夹在两块建设用地边缘,所提出的方法依然可以分割以及提取每一块土地.表 3 为土地分割的均值平均精度 mAP 和轮廓提取的 mADE 评估.结果表明在土地

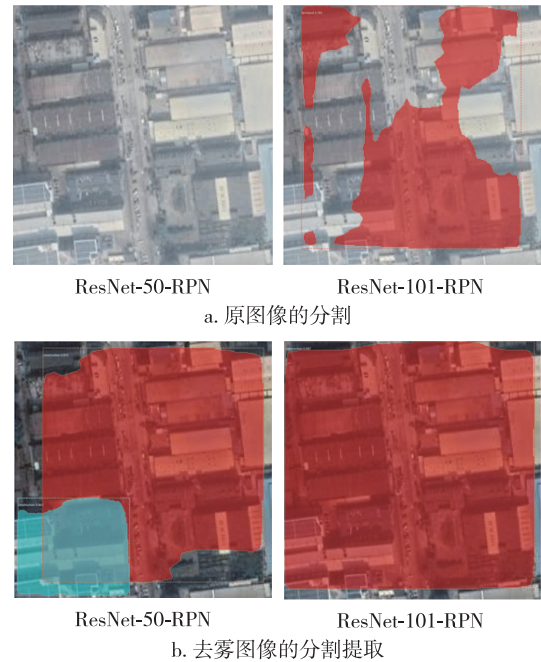


图 4 土地分割与轮廓提取结果在原图像和去雾图像上的对比

Fig. 4 Land segmentation and contour extraction results on original images (a) and defogged images (b)

分割精度上,骨干网络为 ResNet-101-RPN 在去雾图像数据集获取的 mAP 值比原图像数据集高出 3.7 个百分点;在土地轮廓提取精度中,骨干网络为 Res-

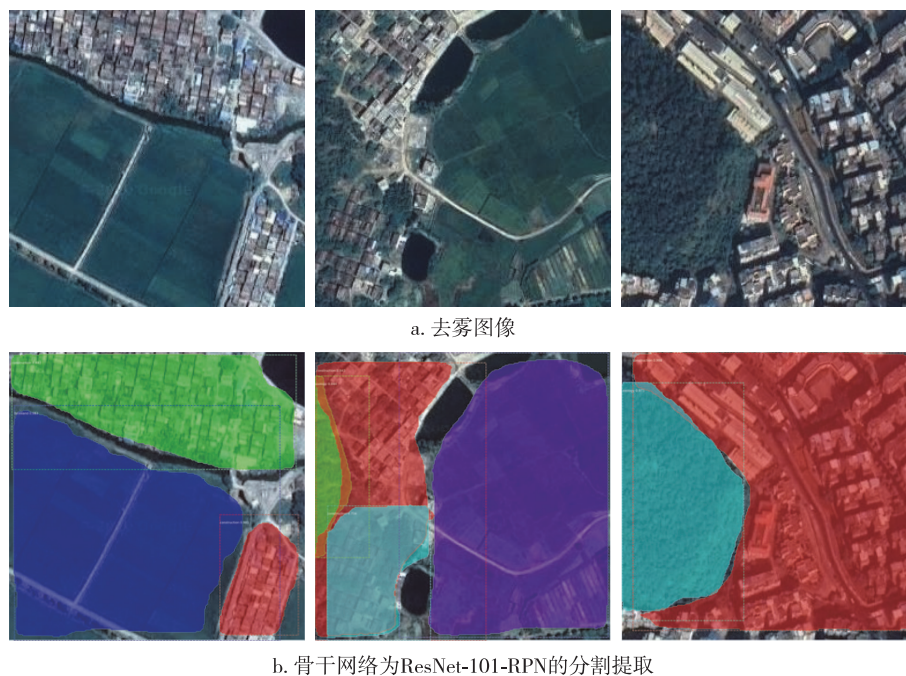


图 5 基于骨干网络为 ResNet-101-RPN 的 Mask R-CNN 在去雾图像上的分割与轮廓提取  
Fig. 5 Land segmentation and contour extraction on defogged images using Mask R-CNN with Resnet-101-RPN as backbone network

Net-101-RPN 在去雾图像数据集的 mADE 为 31.33 像素,优于原图像数据集的 34.67 像素.结果表明,在去雾图像上,改进后的图像质量有助于土地分割与轮廓提取,同时骨干网络 ResNet-101-RPN 提取土地特征的效果比 ResNet-50-RPN 更优.综上所述,在对遥感影像进行土地统计分析时,本文所提出的基于骨干网络为 ResNet-101-RPN 的 Mask R-CNN 在去雾图像上能对土地进行精确分割和轮廓提取.

表 3 土地分割的均值平均精度和轮廓提取的均值平均距离误差

Table 3 Mean average precision for land segmentation and mean average distance error for contour extraction

骨干网络	数据类型	mAP	mADE/像素
ResNet-50-RPN	原图像	0.817	40.52
	去雾图像	0.879	35.21
ResNet-101-RPN	原图像	0.870	34.67
	去雾图像	0.907	31.33

## 5 结论

在国家城镇规划中,高精度的土地分割和轮廓提取是遥感影像进行土地统计分析的基础.为了在不清晰的地图瓦片上实现土地的分割和轮廓提取,本文提出了一种基于骨干为 ResNet-101-RPN 的 Mask R-CNN 的土地分割和轮廓提取方法.首先利用 Google Maps 瓦片方案选用数据集,采用文献[17]算法进行去雾处理,通过降低大气中雾霾细小颗粒、水蒸气等杂质影响来提高图像的质量,进而实现高精度的土地分割和轮廓提取.实验结果表明,基于骨干网络为 ResNet-101-RPN 的 Mask R-CNN 在去雾图像数据集的 mAP 为 0.907,比原图像数据集提高了 3.7 个百分点,同时土地轮廓提取的评估指标 mADE 为 31.33 像素.因此,在遥感影像土地统计分析中,本文提出的方法能够实现高精度的土地分割和轮廓提取.以后将通过改进网络结构来提高重叠土地区域的分割和提取能力.

## 参考文献

### References

- [1] Roy D P, Wulder M A, Loveland T R, et al. Landsat-8: science and product vision for terrestrial global change research[J]. Remote Sensing of Environment, 2014, 145: 154-172
- [2] Geusebroek J, Smeulders A W M, van de Weijer J. Fast anisotropic Gauss filtering [J]. IEEE Transactions on Image Processing, 2003, 12(8): 938-943
- [3] Qu Y D, Cui C S, Chen S B, et al. A fast subpixel edge detection method using Sobel-Zernike moments operator [J]. Image and Vision Computing, 2005, 23(1): 11-17
- [4] Ding L J, Goshtasby A. On the Canny edge detector [J]. Pattern Recognition, 2001, 34(3): 721-725
- [5] 陈志强, 郭永亮, 陈诗哲. 基于小波变换与改进的 Roberts 算子融合的图像边缘检测 [J]. 传感器世界, 2011, 17(4): 15-18  
CHEN Zhiqiang, GUO Yongliang, CHEN Shizhe. Image edge detection based on wavelet transform and improved Roberts operator [J]. Sensor World, 2011, 17(4): 15-18
- [6] Hay G J, Blaschke T, Marceau D J, et al. A comparison of three image-object methods for the multiscale analysis of landscape structure [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2003, 57(5): 327-345
- [7] 彭志平, 魏蛟龙. 一种快速生成面向对象模型的方法 [J]. 计算机工程与应用, 2001, 37(21): 84-86  
PENG Zhiping, WEI Jiaolong. An rapid method of creating object-oriented data model [J]. Computer Engineering and Applications, 2001, 37(21): 84-86
- [8] Vancoillie F, Verbeke L, Dewulf R. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium [J]. Remote Sensing of Environment, 2007, 110(4): 476-487
- [9] Qiao Y, Cappelle C, Ruichek Y, et al. ConvNet and LSH-based visual localization using localized sequence matching [J]. Sensors, 2019, 19(11): 2439
- [10] Kumar S, Pandey A, Sai Ram Satwik K, et al. Deep learning framework for recognition of cattle using muzzle point image pattern [J]. Measurement, 2018, 116: 1-17
- [11] Pinheiro P O, Collobert R, Dollár P. Learning to segment object candidates [C] // Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015, 2: 1990-1998
- [12] Andrew W, Greatwood C, Burghardt T. Visual localisation and individual identification of Holstein Friesian cattle via deep learning [C] // IEEE International Conference of Computer Vision Workshop, 2017: 22-29
- [13] Pinheiro P O, Lin T Y, Collobert R, et al. Learning to refine object segments [M]. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 75-91
- [14] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149
- [15] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651
- [16] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397
- [17] He K M, Sun J, Tang X O. Single image haze removal using dark channel prior [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33

- (12):2341-2353
- [18] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:2117-2125
- [19] Hariharan B, Arbeláez P, Girshick R, et al. Simultaneous detection and segmentation [M]. Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 297-312
- [20] Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning [C] // Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, 2016, 16:265-283
- [21] Russell B C, Torralba A, Murphy K P, et al. LabelMe: a database and web-based tool for image annotation [J]. International Journal of Computer Vision, 2008, 77 ( 1/2/3 ):157-173
- [22] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [M]. Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 740-755

## Land segmentation and contour extraction of remote sensing image based on Mask R-CNN

LI Jiang<sup>1</sup> XU Minghui<sup>1</sup> ZHANG Yu<sup>2</sup>

1 School of Automation, Guangdong University of Technology, Guangzhou 510006

2 Jiangmen Power Supply Bureau, Guangdong Power Grid Company Limited, Jiangmen 529000

**Abstract** Urban spatial planning mainly relies on multiband remote sensing images. However, it is difficult for traditional pattern recognition methods to accurately analyze land covers from remote sensing images. On the premise of accurate land type classification, we propose a method of remote sensing image segmentation and contour extraction based on Mask R-CNN with ResNet-101-RPN as backbone network. The method includes the following steps: data acquisition, image defogging, statistical analysis of land covers on remote sensing image, land segmentation and contour acquisition. The proposed method is trained and tested on a challenging satellite tile dataset. Experimental results show that the method can obtain satisfactory land segmentation and contour extraction results by 0.907 of mean average precision (mAP) and 31.33 pixel of mean average distance error (mADE).

**Key words** land classification; contour extraction; map tiles; Mask R-CNN