

结合 RGB-DSM 图像和深度学习的 城市樟树树冠检测

王昊, 夏凯, 杨垠晖, 冯海林

1. 浙江农林大学 数学与计算机科学学院, 杭州 311300;
2. 浙江省林业智能监测与信息技术研究重点实验室, 杭州 311300;
3. 林业感知技术与智能装备国家林业局重点实验室, 杭州 311300

摘要: 当前, 结合遥感图像和深度学习进行单木树冠检测已经成为一种趋势。RGB 图像是检测中最常用的数据类型, 但由于树冠的颜色和纹理一般比较接近, 在树冠密度较高的区域, 仅使用 RGB 图像的颜色和纹理信息难以区分不同个体的树冠。对此, 本研究在 RGB 图像的基础上, 叠加了高程信息, 以提高单木树冠检测的精度。实验采用彩色图像 RGB 图像和数字表面模型 DSM (Digital Surface Model) 作为数据源, 并分别利用波段组合和双源检测网络模型两种方法结合 RGB 和 DSM 进行单木树冠检测。在前一种方法中, 对 RGB 和 DSM 进行波段组合, 生成 GBD、RGD 和 RBD 3 类图像, 并使用这 3 类图像分别进行网络的训练和测试。在后一种方法中, 将 RGB 和 DSM 输入双源检测网络模型, 并得到检测结果。本文使用 FPN-Faster-R-CNN 和 YOLOv3 进行实验, 相比于 RGB 方案 (仅使用地物的颜色和纹理信息进行单木树冠检测, 是对照方案), FPN-Faster-R-CNN 在 GBD 方案、RBD 方案和双源检测网络方案中的平均精度分别上升了 3.36%、2.45% 和 7.77%, 在 RGD 方案中的平均精度下降了 0.17%, YOLOv3 在 GBD 方案、RBD 方案和双源检测网络方案中的平均精度分别上升了 0.72%、0.14% 和 5.71%, 在 RGD 方案中的平均精度下降了 0.98%。在两个网络下, 双源检测网络方案都在各方案中取得了最佳的检测结果。并且相对于 RGB 方案, 双源检测网络方案在平均精度上的提升幅度随着树冠密度的上升呈现出上升的趋势。对比分析实验结果可知, 在基于深度学习的城市单木树冠检测任务中, 妥善结合并利用地物的颜色、纹理信息和高程信息有利于提高任务性能。

关键词: 遥感, 单木树冠检测, 深度学习, 城市, 高程, 彩色图像, 无人机

中图分类号: TP701/P2

引用格式: 王昊, 夏凯, 杨垠晖, 冯海林. 2023. 结合 RGB-DSM 图像和深度学习的城市樟树树冠检测. 遥感学报, 27(12): 2762-2773
Wang H, Xia K, Yang Y H and Feng H L. 2023. Urban cinnamomum camphora crown detection research using RGB-DSM images and deep learning. National Remote Sensing Bulletin, 27 (12): 2762-2773 [DOI: 10.11834/jrs.20221613]

1 引言

城市森林具有调节城市气候、吸收有毒有害气体、改善居住环境和维持生物多样性等重要作用。随着城市化效应日趋严重, 城市森林的建设愈发受到重视。单木树冠检测 ITCD (Individual Tree Crown Detection) 作为实现城市森林可持续经营和管理的重要方法 (Goldbergs 等, 2018), 不但被用于获取树木的基本信息, 还在病树检测 (Endreny,

2018) 和城市绿量监测 (兰玉彬 等, 2019) 等方面有重要应用。

ITCD 任务可以分为两个步骤: 位置检测和边缘描绘。位置检测的常用算法包括局部最大值法 (Xu 等, 2021b)、模板匹配法、尺度分析法等; 边缘描绘的常用算法包括区域生长法 (Torresan 等, 2020)、分水岭分割法 (Huang 等, 2018)、谷底跟踪法等。在之前很长一段时间, ITCD 任务都是基

收稿日期: 2021-10-22; 预印本: 2022-05-12

基金项目: 国家自然科学基金(编号:U1809208); 浙江省自然科学基金委员会—青山湖科技城管委会联合基金项目(编号:LQY18C160002); 浙江省自然科学基金(编号:LQ20F020005)

第一作者简介: 王昊, 研究方向为林业信息化。E-mail: 524285019@qq.com

通信作者简介: 夏凯, 研究方向为机器学习, 遥感及林业信息化。E-mail: xiakai@zafu.edu.cn

于这些非深度学习的机器学习算法进行的。

随着计算机算力的提升和数据量的增加, 强大的特征学习能力使得深度学习在适应复杂背景、泛化模型和降低准入门槛等方面具备了明显的优势 (Weinstein 等, 2019), 这些优势使得深度学习在人脸识别 (Gong 和 Niu, 2020)、无人驾驶 (Wang, 2020) 等多个领域得到了广泛的应用。在林业上, 结合深度学习和无人机高分影像, ITCD 任务被应用于诸多场景: 为了监测森林虫害, 孙钰等 (2018) 基于油松树冠的无人机航拍彩色图像, 利用 SSD 检测受红脂大小蠹危害的油松。为了快速获取作物信息以优化果园的管理工作, Wu 等 (2020) 使用深度学习检测无人机彩色图像中的苹果树冠。为了获取城市树木的信息, Wegner 等 (2016) 结合使用无人机彩色图像和 Google Map 街景图, 基于深度学习对城市中的 13 类行道树进行检测。为了快速且准确地评估柑橘的产量, Apolo-Apolo 等 (2020) 使用 Faster R-CNN 检测无人机图像中的柑橘树, 并在此基础上进行产量评估模型的训练。

已有研究表明, 高程信息有助于提升计算机视觉任务的性能 (Guo 和 Chen, 2018)。一些研究者在 ITCD 研究中使用了高程数据: Goldbergs 等 (2018) 利用无人机高分影像生成数字表面模型 DSM (Digital Surface Model) 和点云, 进而生成树冠高度模型, 最后结合局部最大值算法和分水岭算法分割出树冠并提取树高等因子。Zhen 等 (2014) 使用激光雷达点云数据生成树冠高度模型, 然后基于正射影像图 DOM (Digital Orthophoto Map) 和树冠高度模型, 利用区域生长法进行树冠边界勾绘。Xu 等 (2021a) 对局部最大值算法进行改进, 提出了一种基于树冠形态信息的方法用于从树冠高度模型中检测树冠。Dong 等 (2020) 提出了一种基于图像分割的方法, 可以从点云数据中提取林下乔木。Nevalainen 等 (2017) 利用无人机彩色图像生成树冠高度模型, 并与高光谱数据结合, 对针叶林进行单木树冠的检测和分类。上述研究在基于非深度学习算法的 ITCD 研究中使用了高程数据, 其采用的这些无监督学习算法更加适合纯林及景观同质化区域 (Xie 等, 2018), 但是本文的研究区位于城市之中, 含有多种类型的植被及其他类型的地物。在这种复杂的环境中, 上述算法难以取得良好的检测效果, 因此本文选择基于深度学习算法进行研究。

目前, 基于深度学习的 ITCD 研究主要使用多光谱图像、高光谱图像、彩色图像作为数据源: Ampatzidis 和 Partel (2019) 使用 YOLOv3 对多光谱图像进行柑橘树冠的检测, 并针对柑橘园中柑橘排列规律这一特点提出了一种数据处理方法和流程, 取得了更优的检测结果。Miyoshi 等 (2020) 提出了一种基于深度学习的方法, 该方法在网络头部加入了波段选择模块, 可以从高光谱数据中选择一定数量的波段, 用以对复杂林地进行单木树冠检测。Torres 等 (2020) 对比了 5 种深度学习方法利用无人机 RGB 影像检测分割单木树冠的能力, 并评估了条件随机场对结果的影响。Dos Santos 等 (2019) 为了检测某一濒危树种, 使用了 3 种经典的深度学习网络对彩色图像进行单木树冠检测。这些研究都依靠遥感图像中地物的颜色和纹理信息检测树冠, 却忽略了地物的高程信息。

在实际应用中, 一些林地的树冠密度较高, 树冠之间难以区分。同时, 树冠本身具有类间差异小, 类内差异大的特点, 在不少情况下, 深度学习网络仅凭借地物的颜色和纹理信息, 难以给出令人满意的检测结果。而高程信息是树冠除了颜色和纹理信息之外, 所具有的另一类型的信息, 也可以用于单木树冠检测。并且, 高程数据易于获取, 可以利用无人机航摄影像和运动结构恢复算法生成。融合地物的颜色、纹理信息和高程信息进行单木树冠检测, 将有助于解决上述问题。

对此, 本研究基于深度学习, 结合 RGB 和 DSM 进行 ITCD 研究, 以探究在城市场景下的 ITCD 任务中, 地物的颜色、纹理信息与高程信息的结合对任务性能的影响。在实验过程中, 由于 RGB 和 DSM 分别是三通道图像和单通道图像 (两者共计有 4 个通道), 而目前通用的深度学习网络接受的是三通道图像, 因此在结合两类数据进行研究之前, 必须设法解决研究数据和网络在通道数量上的矛盾。针对这个问题, 本研究使用了波段组合和双源检测网络模型两种方法进行处理。在前一种方法中, 处理问题的思路是舍弃 RGB 三通道中的一个通道, 然后利用 4 个通道中剩余的 3 个通道构成三通道图像, 进而展开研究。在后一种方法中, 处理问题的思路是对通用的深度学习网络进行改进, 以得到双源检测网络模型, 双源检测网络模型能够同时利用 RGB 图像中包含的颜色和纹理信息以及 DSM 中包含的高程信息, 并不舍弃任意一个通道。

2 研究区概况与数据采集

2.1 研究区概况

研究区位于浙江省杭州市临安区（118°51'E—119°52'E，29°56'N—30°23'N），地处中亚热带季风气候区南缘，夏季炎热湿润，冬季寒冷干燥，春秋两季舒适宜人。由于临安区拥有极其丰富的植物资源，且森林覆盖率达到81.9%，因此被称为“国家园林城市”。本文选取农林大路、武肃街、马溪路、西墅街等街道和浙江农林大学东湖校区内的若干绿化用地和山坡作为数据采集区域。目标区域内含建筑物、道路、路灯、各类交通工具等多类地物，属于典型的城市场景。研究区概况图如图1所示。



图1 研究区概况图
Fig. 1 Overview of the study area

2.2 无人机图像采集

在临安城区中，主要树种有三千余种，这些树木起着净化空气、吸收有毒有害气体、削弱噪音等保健作用。其中，樟树（*Cinnamomum camphora*）是常绿乔木，树冠雄伟，材质优良，作为一种优秀的城乡绿化树种，在研究区内分布得极为广泛，因此本文选择樟树作为研究对象。

本文采用大疆创新科技有限公司的精灵 Phantom 4 RTK 无人机及其搭载的相机作为图像采集系统，

无人机图像采集系统的详细参数见表1。图像于2021年3月3日至2021年3月15日，在天气晴朗或多云、风速较小的环境下，利用无人机自动巡航、定时拍摄的方式获取。本研究共执行飞行任务31次，飞行参数不完全一致。考虑到研究区内的楼高问题，在浙江农林大学校园内执行的飞行任务，航高设为50 m，在浙江农林大学校园外执行的飞行任务，航高设为70 m。为了制作正射影像图，航向重叠和旁向重叠均设为90%。经过上述飞行任务，本研究得到31批，共计11377张无人机航拍图像。

表1 无人机图像采集系统参数
Table 1 UAV image acquisition system parameters

参数		数值
无人机	重量(含桨和电池)	1391 g
	最大水平飞行速度	50 km/h
	最大上升速度	6 m/s
	最大下降速度	3 m/s
	最大可倾斜角度	258°
	飞行时间	约30 min
相机	影像传感器	1英寸CMOS
	像素	2000万
	图像分辨率	5472×3648 pixels
	照片格式	JPEG
视场角		84°

3 研究方法

3.1 目标检测网络

本文使用目标检测网络检测图像中的樟树树冠。其中，目标检测（Xiao等，2020）是指在图像中精准地找到目标物体的位置并判定其类别的过程，在这个过程中使用的卷积神经网络即为目标检测网络。

本文使用FPN-Faster-R-CNN（Lin等，2017）和Yolov3（Redmon和Farhadi，2018）作为研究用的目标检测算法，原因如下：（1）由于环境、树龄、病虫害等因素的影响，树冠大小具有一定的变异性，其中的一些小树冠容易被算法遗漏。由于FPN在小目标检测方面具有良好的效果，因此本文采用FPN-Faster-R-CNN（FPN在Faster-R-CNN上的应用）和Yolov3（FPN在Yolo上的应用）进行研究。（2）FPN-Faster-R-CNN是典型的二阶段网络，Yolov3是典型的一阶段网络，为了使实验结果更加具有广泛性和代表性，本文选择这两种不同类

型的网络进行研究。

3.2 RGB和DSM的结合方法

想要结合RGB和DSM进行ITCD研究，必须解决引言中提到的研究数据和网络在通道数量上的矛盾。针对这个问题，本研究共采用了波段组合和双源检测网络模型两种方法进行处理，具体如下所述。

3.2.1 波段组合

将RGB和DSM进行波段组合的具体做法是：舍弃RGB的3个通道中的其中一个通道，然后利用剩下的两个通道和DSM包含的一个通道构成三通道图像。RGB的3个通道分别代表了研究区的地物在红、绿、蓝3个波段的颜色和纹理信息，记为R、G、B，DSM的一个通道代表了研究区样地的高程信息，记为D。在本研究中，波段组合方案共有3种：对G、B、D的3个波段进行组合（后称GBD方案），对R、G、D的3个波段进行组合（后

称RGD方案），对R、B、D的3个波段进行组合（后称RBD方案）。在上述3种方案中，本研究分别使用不同波段组合构成的图像对深度学习网络进行训练和测试，最终得到3种检测结果。

3.2.2 双源检测网络模型

为了在ITCD任务中充分利用RGB和DSM包含的信息，本研究提出了一种双源检测网络模型，该模型在FPN-Faster-RCNN和Yolov3的基础上改进而来，结构如图2所示：RGB和DSM在经过各自的分支进行特征提取后得到两个特征图，分别在两个特征图上生成感兴趣区域，然后对两组感兴趣区域进行分类，进而得到两组候选检测框（即图2中的Candidate Detection Boxes）。接着对这两组候选检测框进行置信度加权和冗余剔除，得到一组最终的检测框（即图2中的Final Detection Boxes）。最后对检测结果进行评价，得到评价指标（即图2中的Precision、Recall和Average Precision，具体定义详见3.5章节）。

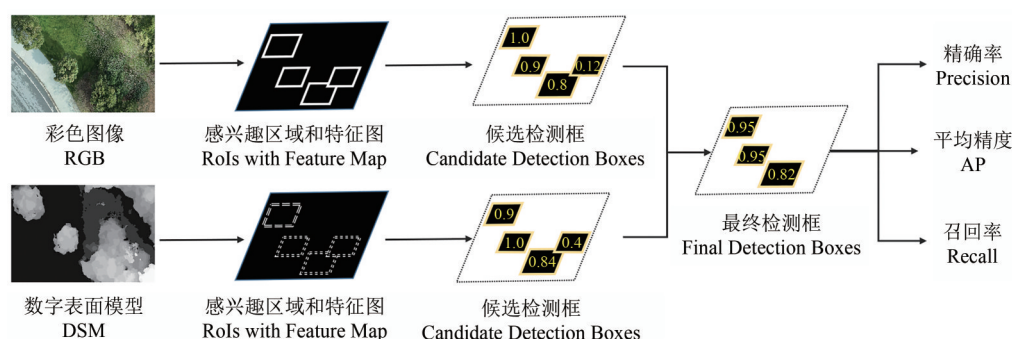


图2 双源检测网络模型(黄色数字代表了对应检测框的置信度。在4个树冠中,最右侧的树冠并非樟树树冠,其对应候选检测框的置信度在加权之后过低,因此在最终检测框中被剔除)

Fig. 2 The double source detection network model (The yellow number represents the confidence of the corresponding detection box. Among the four tree crowns, the rightmost one is not a tree crown of Cinnamomum camphora, and the confidence of its corresponding candidate detection box is too low after weighting, so it is excluded from the final detection box)

其中，对两组候选检测框的具体处理方法如下：若两个检测框分别属于两组候选检测框且代表同一个树冠，则将两者的置信度进行加权，然后赋给原先置信度较高的那个检测框，并剔除另一个检测框；若代表某一物体的检测框只出现在一组候选检测框中，则将其置信度乘以相应的权重。判断两个检测框是否代表同一个树冠的唯一标准是两个检测框之间的交并比IoU (Intersection over Union,)，若IoU大于等于0.5，则认为这两个检测框代表了同一个樟树树冠，若IoU小于0.5，

则认为这两个检测框不是代表同一个樟树树冠。计算公式如式(1)所示：

$$\text{IoU}(X,Y) = \frac{X + Y - X \cup Y}{X \cup Y} \quad (1)$$

式中， X 和 Y 分别代表了两个检测框。

实现上述过程的做法是：将两组候选检测框汇集，其中置信度最高的检测框为 α ，其余的检测框组成一个集合 Boxes ，如式(2)所示。

$$\text{Boxes} = \{\text{box1}, \text{box2}, \dots\} \quad (2)$$

计算检测框 α 与集合 Boxes 中所有检测框的IoU，并取最大的IoU，如式(3)所示。

$$\text{Max_IoU}(\alpha, \text{Boxes}) = \max(\text{IoU}(\alpha, \text{box1}), \text{IoU}(\alpha, \text{box2}), \dots) \quad (3)$$

假设检测框 α 与检测框 β 之间的IoU最大, 式(4)决定了 β 的权重: 如果该IoU大于等于0.5, 则认为 α 和 β 代表了同一个樟树树冠, 给 β 的权重为 weight_β ; 如果该IoU小于0.5, 则认为 α 和 β 不是代表同一个樟树树冠, 给 β 的权重为0。

$$\text{Weight}(\text{max_iou}) = \begin{cases} \text{weight}_\beta, & \text{max_iou} \geq 0.5 \\ 0, & \text{max_iou} < 0.5 \end{cases} \quad (4)$$

最后, 如式(5)所示, 将 α 和 β 进行加权得到一个新置信度。这个新置信度被赋值给检测框 α , 代表在此方案(后称双源检测网络方案)的检测结果中, 检测框 α 里的物体是樟树树冠的概率, 而检测框 β 则被剔除。其中, conf_α 和 conf_β 分别代表 α 和 β 的置信度。重复上述流程, 直到完成所有检测框的处理。

$$\text{Conf}(\alpha, \text{Boxes}) = \text{weight}_\alpha \cdot \text{conf}_\alpha + \text{Weight}(\text{Max_IoU}(\alpha, \text{Boxes})) \cdot \text{conf}_\beta \quad (5)$$

式(4)和式(5)中的 weight_α 和 weight_β 分别代表了检测框 α 和检测框 β 的权重, 其值取决于检测框所属的图像和采用的权重策略, 具体情况见表2。本研究共采用了3种权重策略, 分别是平

分法、波段决定法和AP决定法。平分法对属于两类图像的检测框给出相同的权重, 都为50%。波段决定法根据两类图像的通道数量给出相应的权重, 由于RGB含有3个通道, DSM含有一个通道, 因此给RGB和DSM所属检测框的权重分别为75%和25%。AP决定法根据RGB方案和DSM方案在测试集上所得到的平均精度来给出相应的权重, 平均精度越高, 所属方案的检测框就拥有更高的权重。在这里共有4种情况, 每种情况都基于图3和图4(由于不同权重策略下, 双源检测网络方案的表现相近, 这里将AP决定法下的实验结果作为代表, 图4同。)中RGB方案和DSM方案在两个网络中的平均精度进行计算后给出具体数值。

表2 RGB和DSM的权重配置

Table 2 Weight configuration for RGB and DSM

权重策略	RGB	DSM
平分法	FPN-Faster-R-CNN(50%)	FPN-Faster-R-CNN(50%)
	Yolov3(50%)	Yolov3(50%)
波段决定法	FPN-Faster-R-CNN(75%)	FPN-Faster-R-CNN(25%)
	Yolov3(75%)	Yolov3(25%)
AP	FPN-Faster-R-CNN(51.27%)	FPN-Faster-R-CNN(48.73%)
	Yolov3(50.67%)	Yolov3(49.33%)
决定法	FPN-Faster-R-CNN(52.03%)	Yolov3(47.97%)
	Yolov3(49.91%)	FPN-Faster-R-CNN(50.09%)

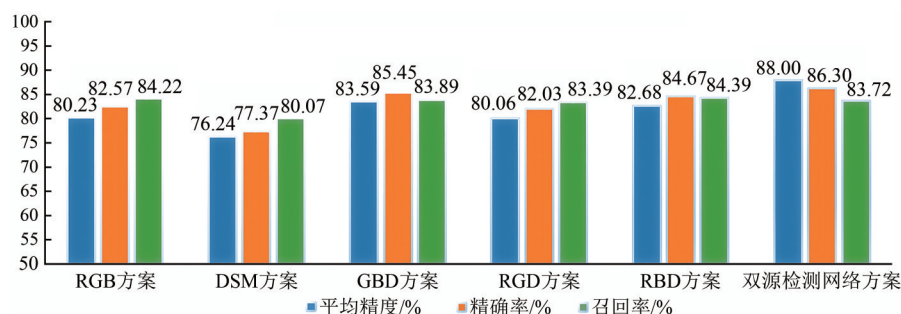


图3 各方案在FPN-Faster-R-CNN下的实验结果

Fig. 3 Experimental results of each scheme under FPN-Faster-R-CNN

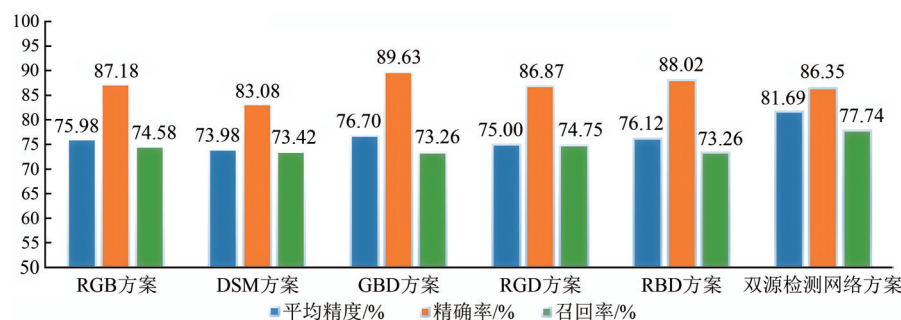


图4 各方案在Yolov3下的实验结果

Fig. 4 Experimental results of each scheme under Yolov3

在对检测结果进行评价时,评价指标的计算依赖于各个检测框的置信度。上述过程中进行的置信度加权这一步骤,便是为了结合利用高程信息和颜色、纹理信息对检测框进行判别,以提高正样本的置信度,降低负样本的置信度,进而改善检测结果。

3.3 数据处理及数据集制作

3.3.1 数据预处理

将无人机航摄任务所得的31批图像分别在 Agisoft Metashape Professional 中加载并生成密集点云,然后制作并导出 RGB (正射影像图) 和 DSM 各31幅。由于一些样地的形状在长宽比上比较极端,因此将 DSM 重采样至与 RGB 相同的分辨率,然后以 1:1 至 2:1 的长宽比裁剪图像,得到一一对应的 RGB 局部图像和 DSM 局部图像各 90 张。

3.3.2 数据集制作

利用 RGB 局部图像和 DSM 局部图像的 4 个通道进行波段组合,得到 GBD 局部图像、RGD 局部图像、RBD 局部图像,再将 RGB、DSM、GBD、RGD 和 RBD 局部图像 (后统称为 5 类图像) 分别划分为训练图像和测试图像。每类图像的构成情况相同,具体如下:测试图像共计 46 张,包含 602 个樟树树冠和 409 个其他树种树冠。训练图像共计 44 张,包含 617 个樟树树冠和 435 个其他树种树冠。为了增加数据量以获得更好的检测效果,对训练图像进行数据增强,通过顺时针 90° 旋转、顺时针 180° 旋转、顺时针 270° 旋转、左右镜像、上下镜像、0.8 倍亮度变换,将训练图像扩充至 308 张,包含 4319 个樟树树冠和 3045 个其他树种树冠。

研究人员通过实地调查确定单木树冠的类别、位置和大小,特别是树冠相互遮盖的地方,是重点调查区域,确保正确记录每一棵树木的信息。在此基础上,使用开源标注工具 Labelimg 完成标注工作。

上述 5 类图像是同源的,因此不用进行影像配准。在进行标注工作的时候,只需要标注其中一类图像,所得的标注文件可以给 5 类图像共用。在实际操作过程中,由于 RGB 局部图像是彩色图像,符合人类视觉的生理特性,因此对其进行标注。将得到的标注文件分别与这 5 类图像组合,得到 RGB 数据集、DSM 数据集、GBD 数据集、RGD 数

据集和 RBD 数据集 (后统称为 5 类数据集)。

3.4 网络训练

本研究使用的计算机操作系统为 Windows 10 (64 位,教育版),CPU 为 Intel Xeon E3-1225 v5@ 3.30 GHz, GPU 为 Nvidia GeForce GTX 1080 Ti, 显存为 11 G。编程语言为 Python, 对应环境版本为 3.5.2。为了加快模型的收敛速度,还分别下载两个网络需要的预训练模型并进行迁移学习,网络的详细参数见表 3。对于两个网络,分别使用 5 类数据集训练网络,直至网络收敛,得到 5 个模型。

表 3 网络详细参数

Table 3 Network detail parameters

参量	FPN-Faster-R-CNN	Yolov3
优化器	标准动量优化器	标准动量优化器
优化器参数	学习率:0.001	学习率:0.001
	动量:0.9	动量:0.9
实现框架	Tensorflow	Keras
主干网络	ResNet101	DarkNet53
batch size	1	32

3.5 评价指标

本研究采用召回率 (Recall)、精确率 (Precision) 和平均精度 AP (Average Precision) 作为评价指标,依据这 3 个指标在测试集上的表现评价一个模型的优劣。在本研究中,召回率表示实际存在的樟树树冠有多少被网络检测到,精确率表示网络检测到的樟树树冠有多少是正确的。在不同的置信度下,精确率和召回率不同,由此可以绘制出一条 P-R 曲线,平均精度在数值上等于 P-R 曲线下的面积。相关计算公式如下:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(6)

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

(7)

$$\text{AP} = \int_0^1 p(r)dr$$

(8)

式中, TP (True Positive) 表示被判定为樟树树冠,实际上也是樟树树冠的检测框。FP (False Positive) 表示被判定为樟树树冠,但实际上不是樟树树冠的检测框。FN (False Negative) 表示没有被判定为樟树树冠,但实际上是樟树树冠的检测框。具体的判定方法如下:取图像对应的标注文件中所有的矩形标注框,一一计算检测框与这些标注框

的IoU。取最大的IoU及其对应的标注框，若该IoU大于等于0.5，则认为检测框和标注框所代表的樟树树冠是同一个，检测正确；若该IoU小于0.5，则认为检测框所在的位置不存在樟树树冠，检测错误。

4 实验结果与分析

4.1 各方案检测结果的对比分析

本研究共采用了两个网络（每个网络6种方案）进行ITCD研究，各方案的检测结果如图3和图4所示，表4详细展示了双源检测网络在各种情况下的检测结果。其中，精确率和召回率是置信度大于等于置信度阈值的检测框而得的，经过实验，Yolov3、FPN-Faster-R-CNN、同时使用Yolov3和FPN-Faster-R-CNN的双源检测网络的置信度阈值分别设为0.3、0.5和0.5。图5展示了各方案的检测结果示例图，其中，双源检测网络方案只展示了最佳权重策略（AP决定法）的结果图。

表4 双源检测网络方案在不同情况下的检测结果
Table 4 Detection results of double source detection network scheme under different conditions

方法	平均精度	精确率	召回率
	AP/%	Precision/%	Recall/%
FPN-RGB FPN-DSM 平分法	87.76	86.12	85.55
FPN-RGB FPN-DSM 波段决定法	87.93	83.61	83.89
FPN-RGB FPN-DSM AP决定法	88.00	86.30	83.72
Yolo-RGB Yolo-DSM 平分法	81.68	86.19	77.74
Yolo-RGB Yolo-DSM 波段决定法	81.56	88.36	74.42
Yolo-RGB Yolo-DSM AP决定法	81.69	86.35	77.74
FPN-DSM Yolo-RGB 平分法	84.95	89.45	73.26
FPN-DSM Yolo-RGB 波段决定法	85.52	93.51	71.76
FPN-DSM Yolo-RGB AP决定法	84.95	86.50	75.58
FPN-RGB Yolo-DSM 平分法	87.58	90.18	82.39
FPN-RGB Yolo-DSM 波段决定法	87.59	84.08	84.22
FPN-RGB Yolo-DSM AP决定法	87.44	86.13	83.55

RGB方案是对照方案，FPN-Faster-R-CNN和Yolov3的平均精度分别达到了80.23%和75.98%，整体表现在6种方案中处于中下游。

DSM方案是表现最差的方案，相比于其他5个方案，其检测结果出现了更多的漏判和错判。这说明在ITCD领域，仅凭高程信息不足以获得理想的检测效果，可能的原因是：在树冠非常密集的区域，仅凭借高程信息，网络容易遗漏一些树冠，一些在树形上与樟树相似的树种也容易使网络造成误判。

双源检测网络方案取得了最佳结果，相比于

RGB方案，FPN-Faster-R-CNN下的双源检测网络和Yolov3下的双源检测网络，平均精度分别提高了7.77%和5.71%，且其精确率和召回率也相对靠前。而两个分支同时使用了FPN-Faster-R-CNN和Yolov3的双源检测网络，其平均精度介于上述两者之间。以上实验结果说明，双源检测网络方案有效结合并利用了遥感影像中的高程信息和颜色、纹理信息，提高了ITCD任务的性能。如图6所示，RGB方案漏判了下方的两个樟树树冠，DSM方案漏判了上方的一个樟树树冠，而在双源检测网络方案中，3个树冠都被正确检测。此外，在不同的权重策略下，各个双源检测网络的平均精度波动都控制在0.6%之内，这说明本研究采用的3种权重策略具有相近且良好的效果。

相比于RGB方案，两个网络在GBD方案和RBD方案上的平均精度都有所提高，而在RGD方案上的平均精度都有所下降。这说明虽然红色波段或者绿色波段的缺失损失了一部分颜色信息，但是高程信息带来的增益不但抵消了这部分损失，还提高了检测精度。而蓝色波段的缺失所造成的损失则是略大于高程信息带来的增益，可能的原因是：叶绿素、叶红素和叶黄素都在0.45 μm附近（蓝色波段）有一个吸收带，这使得蓝色波段所包含的颜色信息对ITCD任务而言比较重要。

在平均精度上，双源检测网络方案比进行波段组合的3个方案都要高，这说明：RGB和DSM的4个通道都包含了有助于识别樟树树冠的特征，任意一个通道的缺失都会造成损失，同时利用所有通道才是最有利于ITCD任务的做法。

4.2 不同疏密度下检测结果的对比分析

双源检测网络方案在各个方案中取得了最佳的检测结果，其平均精度相对于RGB方案取得了不小的提升。为了进一步探究该提升的幅度与林地树冠密度之间的关系，本研究取测试集中部分典型样地，将其分为低疏密度、中疏密度和高疏密度3类，并观察这3类林地实验中的表现，以做出定性判断。

其中，对疏密度的分类办法如下：对图像中的每个标注框，一一计算该标注框与剩余树冠标注框间的IoU，然后求和，并对每个标注框的IoU之和取平均值，作为该图像的平均IoU。若平均IoU小于等于0.05，则将该图像归属于低疏密度；若大于0.05且小于等于0.15，则归属于中疏密度；

若大于0.15，则归属于高疏密度。在低疏密度的图像中，树冠之间大多彼此分离且保持了一定的距离；在中疏密度的图像中，树冠大多相互靠近

以至于相切或者有轻微重叠；在高疏密度的图像中，树冠大多有较高的重叠，部分区域甚至难以通过目视解译区分不同的单木。

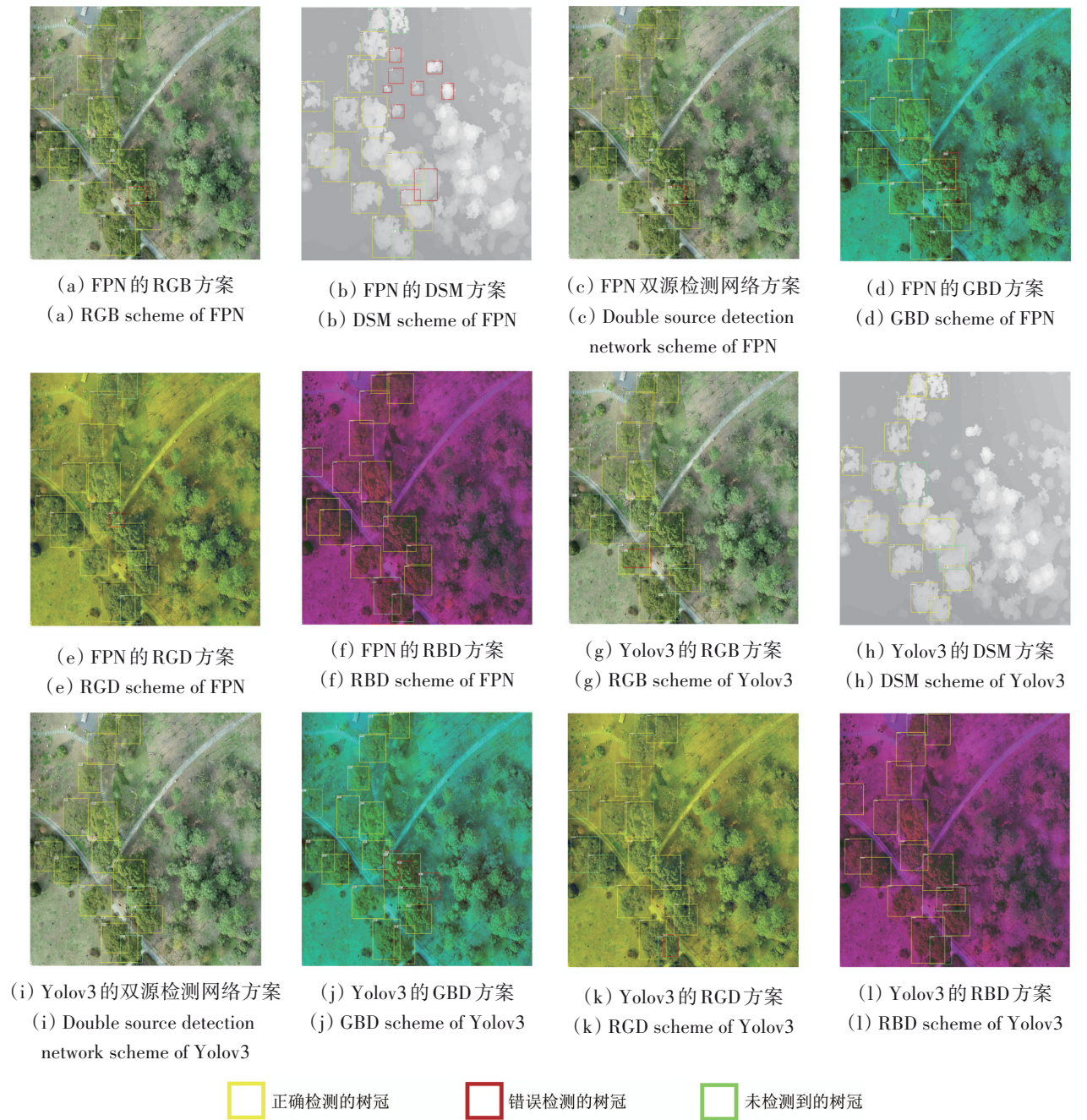


图5 检测结果对比

Fig.5 Comparison of detection images

实验结果如图7和图8所示，随着疏密度的上升，双源检测网络对任务性能的增益也在扩大：同样使用FPN-Faster-R-CNN，在低、中、高3种疏密度的情况下，双源检测网络方案相对于RGB方案在平均精度上分别提升了约3.4%、5.8%、11.7%；同样使用Yolov3，在低、中、高3种疏密度的情况

下，双源检测网络方案相对于RGB方案在平均精度上分别提升了约0.4%、3.7%、6.4%。出现这种趋势的可能原因是：在低疏密度的情况下，树冠与树冠之间保持一定的距离，神经网络凭借地物的颜色和纹理信息就足以精准地检测出樟树树冠，在这种情况下，所能提升的空间相对较小，所以高

程信息的加入仅能为任务性能带来相对较小的增益；而随着疏密度的增加，树冠与树冠之间的重叠率越来越高，神经网络仅凭借地物的颜色和纹理

信息容易出现更多的漏检和错检，在这种情况下，高程信息中所包含的树冠特征具有更大的发挥空间，可以减少这些漏检和错检情况的出现。

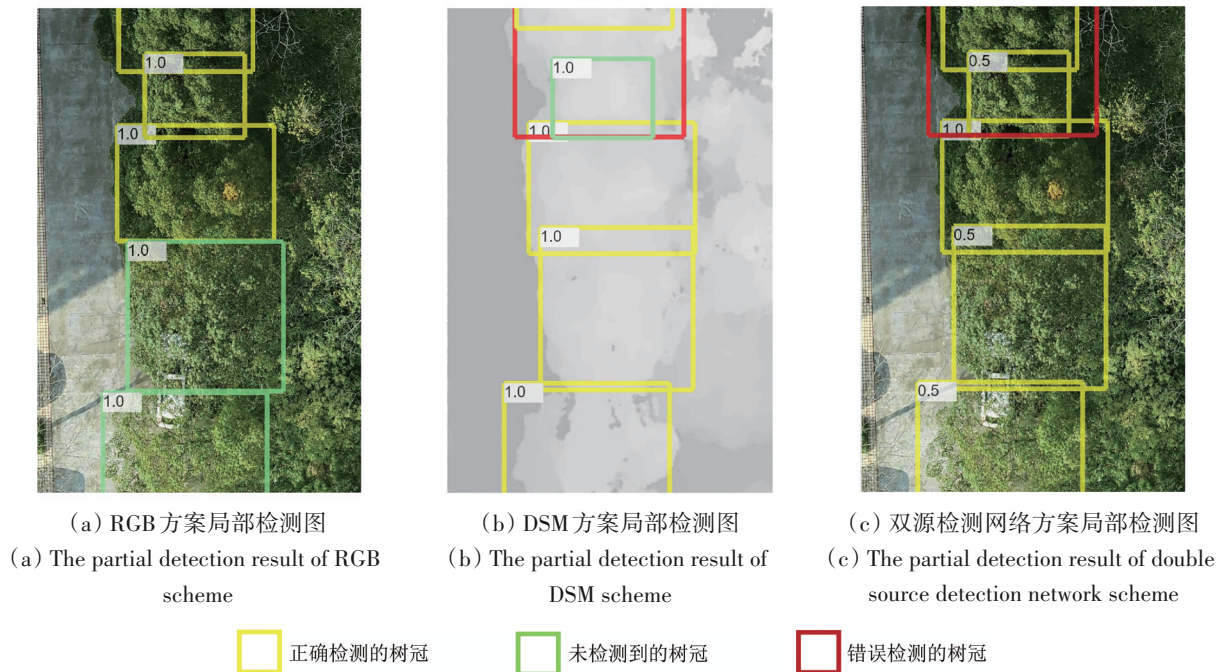


图6 局部检测图像对比

Fig.6 Comparison of partial detection images

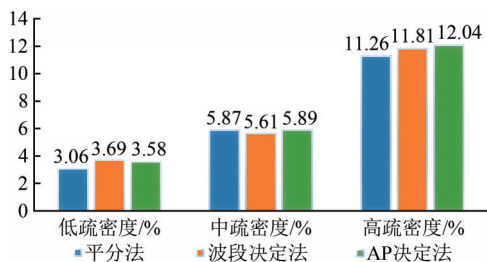


图7 FPN-Faster-R-CNN下的双源检测网络方案相对于RGB方案的平均精度提升幅度

Fig. 7 Average accuracy improvement of double source detection network scheme compared with RGB scheme under FPN-Faster-R-CNN

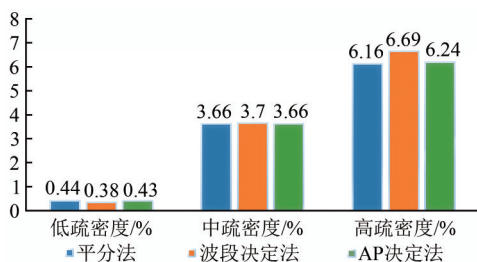


图8 YOLOv3下双源检测网络方案相对于RGB方案的平均精度提升幅度

Fig. 8 Average accuracy improvement of double source detection network scheme compared with RGB scheme under YOLOv3

4.3 樟树树冠检测的误差分析

ITCD任务中可能出现的误差可以总结为以下几类：(1) 漏检，即图像中存在的某一樟树树冠没有被检测出来。(2) 错检，即错误地将其他树种的树冠或者非树冠物体识别为樟树树冠。(3) 过检，即某一樟树树冠被检测为若干个樟树树冠。图9展示了上述错误情况：图9(a)图中，两个较小的樟树树冠被检测为一个较大的樟树树冠，进而导致两个樟树树冠被漏检。图9(b)图中，两个樟树树冠直接被漏检。图9(c)图中，一个桂花树冠被错检为樟树树冠。图9(d)图中，一个樟树树冠被过检为两个樟树树冠。在本研究出现的错误情况中，错检的情况最多，漏检的情况相对少一些，过检的情况极少。

错检情况的出现主要有两方面原因：一方面，樟树树冠会随着树龄、树叶稀疏度、开花与否、修剪与否、光照条件等因素的变化在图像上呈现出一定程度的差异性，这是RGB方案出现错检的主要原因；另一方面，不少树种在树形上与樟树相似，这是DSM方案出现错检的主要原因。由于

进行了加权，双源检测网络方案会同时出现 DSM 方案和 RGB 方案的错检情况，但是加权后的置信度一般较低。而少量的漏检情况主要是由以下两个原因造成的：第一个原因是某些地方的树冠非常密集，在彼此遮盖之下，树冠的细节无法在图像上充分展示，这使得网络难以对树冠进行区分。这些地方主要是一些长时间未对树冠进行修剪的绿化用地，树冠在彼此簇拥之下即使是肉眼也难以分辨。第二个原因是某些单木树冠较小或者树叶比较稀疏，这使得该单木树冠在特征图中所占的像素较少或者表现出的特征不够明显。极少数的过检情况也是由于单木树冠过于密集造成的。

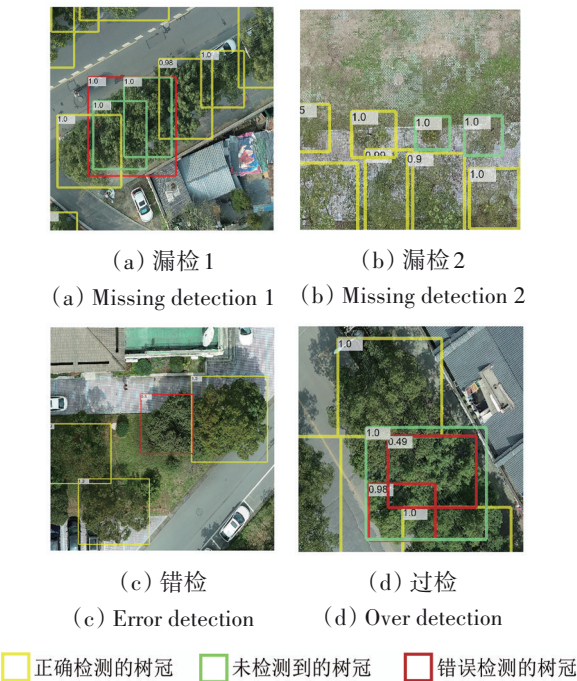


图9 典型的单木树冠检测误差局部图

Fig. 9 Typical ITCD error

5 结 论

当前基于深度学习的 ITCD 研究依靠遥感图像中的颜色和纹理信息来检测树冠，却忽略了高程信息。高程信息也反映出了树冠的特征，所以本研究基于深度学习，并通过波段组合和双源检测网络模型两种方法，结合代表颜色和纹理信息的 RGB 图像和代表高程信息的 DSM 进行 ITCD 研究，以期改善任务性能。

由实验结果可知，双源检测网络方案取得了最佳的检测结果。相比于 RGB 方案的平均精度，FPN-

Faster-R-CNN 下的双源检测网络方案和 YOLOv3 下的双源检测网络方案分别提高了 7.77% 和 5.71%，而两个分支同时使用 FPN-Faster-R-CNN 和 YOLOv3 的双源检测网络方案则介于上述两者之间。在 RGB 方案与 DSM 方案的平均精度相差不大的情况下，本文采用的 3 种权重策略具有相近的效果，双源检测网络方案在这 3 种权重策略下所取得的平均精度仅有微小的波动。此外，双源检测网络方案对任务性能的改善幅度受到林地树冠密度的影响，随着密度的升高，双源检测网络方案对平均精度的提升幅度呈现出上升的趋势。而波段组合的 3 个方案中，RBD 方案和 GBD 方案在两个网络上都取得了比 RGB 方案更好的检测结果，但是 RGD 方案在两个网络上的表现则都比 RGB 方案差。分析本文的实验结果可知，在基于深度学习的城市 ITCD 任务中，妥善地结合并利用 RGB 图像和 DSM 有利于任务性能的提高。而波段组合方法中的 RGD 方案的表现比 RGB 方案差，可能是由于缺失了包含多种色素吸收带的蓝色波段。对比各方案的实验结果，同时使用所有的通道对 ITCD 任务才最有利。

单木树冠检测过程中，误差来源有 3 种：错检、漏检和过检。其中，出现错检的原因是同类树冠间的差异性和不同类树冠间的相似性，出现漏检的原因是树冠间的重叠率过高、树冠过小和树冠的树叶过于稀疏，而出现过检的原因也是树冠间的重叠率过高。

目前，本研究仅在城市这一场景下讨论了 RGB 图像和 DSM 的结合对 ITCD 任务性能的影响，下一步的任务可以考虑在更多的场景下进行实验。

参考文献 (References)

Ampatzidis Y and Partel V. 2019. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*, 11(4): 410 [DOI: 10.3390/rs11040410]
Apolo-Apolo O E, Martínez-Guanter J, Egea G, Raja P and Pérez-Ruiz M. 2020. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. *European Journal of Agronomy*, 115: 126030 [DOI: 10.1016/j.eja.2020.126030]
Dong T Y, Zhang X P, Ding Z F and Fan J. 2020. Multi-layered tree crown extraction from LiDAR data using graph-based segmentation. *Computers and Electronics in Agriculture*, 170: 105213 [DOI: 10.1016/j.compag.2020.105213]
Dos Santos A A, Junior J M, Araújo M S, Di Martini D R, Tetila E C, Siqueira H L, Aoki C, Eltner A, Matsubara E T, Pistori H, Feitosa

- R Q, Liesenberg V and Gonçalves W N. 2019. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors*, 19(16): 3595 [DOI: 10.3390/s19163595]
- Endreny T A. 2018. Strategically growing the urban forest will improve our world. *Nature Communications*, 9(1): 1160 [DOI: 10.1038/s41467-018-03622-0]
- Goldbergs G, Maier S W, Levick S R and Edwards A. 2018. Efficiency of individual tree detection approaches based on light-weight and low-cost UAS imagery in Australian savannas. *Remote Sensing*, 10(2): 161 [DOI: 10.3390/rs10020161]
- Gong T Y and Niu H Q. 2020. An implementation of ResNet on the classification of RGB-D images//2nd BenchCouncil International Symposium on Benchmarking, Measuring, and Optimizing. Denver: 149-155 [DOI: 10.1007/978-3-030-49556-5_15]
- Guo Y R and Chen T. 2018. Semantic segmentation of RGBD images based on deep depth regression. *Pattern Recognition Letters*, 109: 55-64 [DOI: 10.1016/j.patrec.2017.08.026]
- Huang H Y, Li X and Chen C C. 2018. Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(7): 2253-2262 [DOI: 10.1109/JSTARS.2018.2830410]
- Lan Y B, Zhu Z H, Deng X L, Lian B Z, Huang J Y, Huang Z X and Hu J. 2019. Monitoring and classification of citrus Huanglong-bing based on UAV hyperspectral remote sensing. *Transactions of the Chinese Society of Agricultural Engineering*, 35(3): 92-100 (兰玉彬, 朱梓豪, 邓小玲, 练碧桢, 黄敬易, 黄梓效, 胡洁. 2019. 基于无人机高光谱遥感的柑橘黄龙病植株的监测与分类. *农业工程学报*, 35(3): 92-100) [DOI: 10.11975/j.issn.1002-6819.2019.03.012]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of the 2017 IEEE Conference on Computer Vision and pattern Recognition. Honolulu: IEEE: 936-944 [DOI: 10.1109/cvpr.2017.106]
- Miyoshi G T, dos Santos Arruda M, Osco L P, Junior J M, Gonçalves D N, Imai N N, Tommaselli A M G, Honkavaara E and Gonçalves W N. 2020. A novel deep learning method to identify single tree species in UAV-based hyperspectral images. *Remote Sensing*, 12(8): 1294 [DOI: 10.3390/rs12081294]
- Nevalainen O, Honkavaara E, Tuominen S, Viljanen N, Hakala T, Yu X W, Hyypä J, Saari H, Pölönen I, Imai N N and Tommaselli A M G. 2017. Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. *Remote Sensing*, 9(3): 185 [DOI: 10.3390/rs9030185]
- Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement. *arXiv preprint arXiv: 1804.02767* [DOI: 10.48550/arXiv.1804.02767]
- Sun Y, Zhou Y, Yuan M S, Liu W P, Luo Y Q and Zong S X. 2018. UAV real-time monitoring for forest pest based on deep learning. *Transactions of the Chinese Society of Agricultural Engineering*, 34(21): 74-81 (孙钰, 周焱, 袁明帅, 刘文萍, 骆有庆, 宗世祥. 2018. 基于深度学习的森林虫害无人机实时监测方法. *农业工程学报*, 34(21): 74-81) [DOI: 10.11975/j.issn.1002-6819.2018.21.009]
- Torres D L, Feitosa R Q, Happ P N, La Rosa L E C, Junior J M, Martins J, Bressan P O, Gonçalves W N and Liesenberg V. 2020. Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. *Sensors*, 20(2): 563 [DOI: 10.3390/s20020563]
- Torresan C, Carotenuto F, Chiavetta U, Miglietta F, Zaldei A and Gioli B. 2020. Individual tree crown segmentation in two-layered dense mixed forests from UAV LiDAR data. *Drones*, 4(2): 10 [DOI: 10.3390/drones4020010]
- Wang W H. 2020. Detection of panoramic vision pedestrian based on deep learning. *Image and Vision Computing*, 103: 103986 [DOI: 10.1016/j.imavis.2020.103986]
- Wegner J D, Branson S, Hall D, Schindler K and Perona P. 2016. Cataloging public objects using aerial and street-level images—urban trees//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 6014-6023 [DOI: 10.1109/CVPR.2016.647]
- Weinstein B G, Marconi S, Bohlman S, Zare A and White E. 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11): 1309 [DOI: 10.3390/rs11111309]
- Wu J T, Yang G J, Yang H, Zhu Y H, Li Z H, Lei L and Zhao C J. 2020. Extracting apple tree crown information from remote imagery using deep learning. *Computers and Electronics in Agriculture*, 174: 105504 [DOI: 10.1016/j.compag.2020.105504]
- Xiao Y Z, Tian Z Q, Yu J C, Zhang Y S, Liu S, Du S Y and Lan X G. 2020. A review of object detection based on deep learning. *Multimedia Tools and Applications*, 79(33/34): 23729-23791 [DOI: 10.1007/s11042-020-08976-6]
- Xie Y Q, Bao H, Shekhar S and Knight J. 2018. A TIMBER framework for mining urban tree inventories using remote sensing datasets//2018 IEEE International Conference on Data Mining (ICDM). Singapore: IEEE: 1344-1349 [DOI: 10.1109/ICDM.2018.00183]
- Xu W B, Deng S S, Liang D and Cheng X J. 2021a. A crown morphology-based approach to individual tree detection in subtropical mixed broadleaf urban forests using UAV LiDAR data. *Remote Sensing*, 13(7): 1278 [DOI: 10.3390/rs13071278]
- Xu X J, Zhou Z S, Tang Y and Qu Y L. 2021b. Individual tree crown detection from high spatial resolution imagery using a revised local maximum filtering. *Remote Sensing of Environment*, 258: 112397 [DOI: 10.1016/j.rse.2021.112397]
- Zhen Z, Quackenbush L J and Zhang L J. 2014. Impact of tree-oriented growth order in marker-controlled region growing for individual tree crown delineation using airborne laser scanner (ALS) data. *Remote Sensing*, 6(1): 555-579 [DOI: 10.3390/rs6010555]

Urban cinnamomum camphora crown detection research using RGB-DSM images and deep learning

WANG Hao, XIA Kai, YANG Yinhui, FENG Hailin

1. College of Mathematics and Computer Science, Zhejiang Agriculture and Forestry University, Hangzhou 311300, China;

2. Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Hangzhou 311300, China;

3. Key Laboratory of Forestry Perception Technology and Intelligent Equipment, State Forestry Administration, Hangzhou 311300, China

Abstract: Currently, combining remote sensing imagery with deep learning is a growing trend in individual tree crown detection. RGB image is the most commonly used data type in detection. However, given that the color and texture of the tree crowns are generally close, distinguishing the crowns of different individuals by using only the color and texture information of RGB image in areas with high density of crowns is difficult. In this study, the elevation information is superimposed to improve the accuracy of individual tree crown detection by using RGB images. In the experiment, RGB image (color image) and DSM (digital surface model) were used as data sources, and band combination and double-source detection network model were used to combine RGB and DSM for individual tree crown detection. In the former method, band combination of RGB and DSM was conducted to generate GBD, RGD, and RBD images, and the three kinds of images were used for network training and testing. In the latter method, RGB and DSM were input into the double-source detection network model, and the detection results were obtained. FPN-Faster-R-CNN and Yolov3 were used for experiments in this study. Compared with RGB scheme as the control scheme (which uses only the color and texture information of ground objects for individual tree crown detection), the average accuracy of FPN-Faster-R-CNN in the GBD scheme, RBD scheme, and double-source detection network scheme increased by 3.36%, 2.45%, and 7.77%, respectively; it decreased by 0.17% in the RGD scheme. The average accuracy of Yolov3 in the GBD scheme, RBD scheme, and double-source detection network scheme increased by 0.72%, 0.14%, and 5.71%, respectively; it decreased by 0.98% in the RGD scheme. Under the two networks, the double-source detection network scheme achieved the best detection result in each scheme. Compared with the RGB scheme, the improvement in average accuracy of double-source detection network scheme showed a rising trend with the increase in forest density. Comparative analysis of the experimental results shows that proper combination and utilization of the color, texture, and elevation information of the ground objects is beneficial to improve the performance in the urban individual tree crown detection task based on deep learning.

Key words: remote sensing, individual tree crown detection, deep learning, urban, elevation, color image, UAV

Supported by National Natural Science Foundation of China (No. U1809208); Natural Science Foundation of Zhejiang Province, China (No. LQY18C160002); Natural Science Foundation of Zhejiang Province, China (No. LQ20F020005)