Hindawi
Journal of Sensors
Volume 2021, Article ID 6668934, 12 pages
https://doi.org/10.1155/2021/6668934

Research Article

Automatic Detection and Parameter Estimation of *Ginkgo biloba* in Urban Environment Based on RGB Images

Kai Xia[1,2,3] **Hao Wang**[1,2,3] **Yinhui Yang**[1,2,3] **Xiaochen Du**[1,2,3] **and Hailin Feng**[1,2,3]

[1]*College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou, China*
[2]*Key Laboratory of State Forestry and Grassland Administration on Forestry Sensing Technology and Intelligent Equipment, Hangzhou, China*
[3]*Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Hangzhou, China*

Correspondence should be addressed to Yinhui Yang; yhyang@zafu.edu.cn and Hailin Feng; zafu_fal@yeah.net

Received 4 November 2020; Accepted 12 July 2021; Published 6 August 2021

Academic Editor: Liu Hongxiao

Individual tree crown detection and morphological parameter estimation can be used to quantify the social, ecological, and landscape value of urban trees, which play increasingly important roles in densely built cities. In this study, a novel architecture based on deep learning was developed to automatically detect tree crowns and estimate crown sizes and tree heights from a set of red-green-blue (RGB) images. The feasibility of the architecture was verified based on high-resolution unmanned aerial vehicle (UAV) images using a neural network called FPN-Faster R-CNN, which is a unified network combining a feature pyramid network (FPN) and a faster region-based convolutional neural network (Faster R-CNN). Among more than 400 tree crowns, including 213 crowns of *Ginkgo biloba*, in 7 complex test scenes, 174 ginkgo tree crowns were correctly identified, yielding a recall level of 0.82. The precision and *F*-score were 0.96 and 0.88, respectively. The mean absolute error (MAE) and mean absolute percentage error (MAPE) of crown width estimation were 0.37 m and 8.71%, respectively. The MAE and MAPE of tree height estimation were 0.68 m and 7.33%, respectively. The results showed that the architecture is practical and can be applied to many complex urban scenes to meet the needs of urban green space inventory management.

1. Introduction

Urban trees play important roles in densely built cities, with activities that include reducing atmospheric carbon dioxide, alleviating the urban heat island effect [1, 2], isolating noise [3], alleviating urban flood risk [4], and providing shelters for wildlife [5, 6]. Detailed data on urban trees, such as species, location, number, diameter at breast height (DBH), tree height, and crown size, are essential for quantifying these benefits. Traditionally, tree attributes are obtained by field measurement, which is labor- and cost-intensive [7]. Individual tree crown detection (ITCD) technology based on remote sensing, which has the advantage of providing spatially explicit data, potentially with fine temporal resolution and low cost [6], can facilitate urban green space inventory development and monitoring.

ITCD technology has traditionally been used in forest monitoring and consists of 2 phases: (1) locating and delineating individual tree crowns and (2) classifying tree species and estimating morphological parameters such as crown size, tree height, and DBH ([8]).

In the location and delineation phase, the data source and computational methods are the two major factors dominating the results. Regardless of the data source, single tree crowns should first be detected automatically by methods such as local maxima [9], template matching [10], and image binarization [11]; then, crown edges should be delineated by various methods, such as region growing [12], watershed segmentation [13], and valley flowing [14]. Data are classified into three categories, i.e., passive sources (i.e., visible light, multispectral, and hyperspectral), active sources (i.e., LIDAR and radar), and both passive and active sources; data type has

a great impact on the results [15]. In recent decades, an increasing number of LIDAR-based ITCD studies have been carried out because LIDAR provides accurate 3D surface information. In addition, crown delineation has been carried out based on multispectral imagery involving wavelength bands crucial for the identification of vegetation characteristics [16].

In phase 2, the tree crown locations and shapes delineated in phase 1 are applied for species classification or parameter estimation. For example, tree crowns extracted from hyperspectral or multispectral images can be classified accurately based on methods such as support vector machine and random forest [6, 17, 18]. Crown size can be obtained by directly measuring tree crown shape. If the species of the tree crowns is known, we can infer some parameters, such as DBH and tree height, based on linear regression models which determine the relationships between crown size and other parameters [7]. If 3D surface information is also available, tree height can be directly extracted [19–21], as we discuss below. LIDAR data were the earliest and most accurate data type for estimating tree height. Morsdorf et al. [22] derived tree height from segmented individual trees based on LIDAR point clouds, and the accuracy evaluation revealed a strong relation between estimated and field-measured tree height. With the subsequent development of structure-from-motion (SFM) technology, 3D point clouds can be generated from remote sensing images, and it became a popular way to estimate tree height due to its lower cost and easier acquisition process than LIDAR technology. According to various studies in which tree height has been estimated based on SFM technology [7, 19, 23–25], high agreement between remote sensing estimation and field measurements can be achieved.

In contrast to forests, the focus in cities is on individual trees rather than forest stands [26]. The study of urban trees by remote sensing faces the following challenges: (1) urban trees are distributed in complex environments with interfering backgrounds, e.g., buildings, lawns, and low vegetation; (2) urban trees are unevenly distributed, vary in size, and are often in groups with heavy canopy overlap [27]; and (3) there may be many tree species in a small area [26]. In recent years, many approaches have been proposed to tackle these challenges using remote sensing data. Lin et al. [28] developed a three-step method applicable for the detection of individual trees in unmanned aerial vehicle (UAV) oblique images. Gomes et al. [29] realized individual urban tree crown detection in submeter satellite imagery using marked point processes and a geometrical-optical model. Liu et al. [6] and Mozgeris et al. [26] identified tree species in high-resolution spectral images after detecting and segmenting individual tree crowns based on LIDAR technologies. These studies indicate that the detection and classification of urban tree crowns are important but remain difficult. Some specialized methods designed according to the characteristics of the city have been proposed to estimate the morphological parameters of urban trees. Wu et al. [30] used mobile laser scanning to extract street tree height, crown size, and DBH. Jiao and Deng [31] estimated tree height based on the size of the tree shadow using sun angle and the time when the image was taken.

Previous studies on forest and urban ITCD have all demonstrated the great value of detecting tree species and estimating important morphological tree parameters accurately and automatically. However, these studies were based on different types of data sources and a variety of methods. The lack of a standardized pipeline for data acquisition and processing and an integrated computational architecture limit the practical application of current methods. Therefore, there is an urgent need to establish an integrated architecture for practical use, the design of which should be focused on solving the challenges of ITCD of urban trees via the following capabilities: (1) accurately detecting specific tree species in urban environments; (2) inferring key morphological tree parameters; (3) performing computations in a fully automatic way; and (4) performing data acquisition and processing to support the computations in a manner that is not only convenient and economical but also easy to standardize.

In recent years, deep learning approaches, especially deep convolution neural networks, have achieved great performance in object detection, and ITCD studies based on deep convolution neural networks have achieved good results. Morales et al. [32] segmented crowns of *Mauritia flexuosa* in the Amazon rainforest based on the DeepLab v3+ architecture with an accuracy of 96.60%. Ampatzidis and Partel detected citrus trees with a precision and recall of 99.9% and 99.7%, respectively, based on deep learning networks [33].

Considering the lack of an integrated architecture for ITCD with practical capabilities, this study proposes an automated urban canopy detection architecture based on deep learning that can obtain the number, location, crown size, and tree height of a given tree species from a set of RGB images. Specifically, a neural network, FPN-Faster R-CNN, is adapted in this study, and rich evaluation experiments are conducted based on high-resolution UAV images. The results indicate that the architecture is promising for ITCD studies of urban trees, although much more work is needed to further improve its performance in diverse real applications.

2. Materials and Methods

2.1. Study Sites. In this research, three study sites in Linan were selected: one on the campus of Zhejiang Agriculture and Forestry University (ZAFU), one on the west shoreline of Qingshanhu Lake, and one in a residential area (Figure 1). Linan, a typical forest city, is located in Hangzhou, Zhejiang Province, China, centered at latitude/longitude 119.72°/30.25°. It features a subtropical monsoon climate, and the vegetation is dominated by subtropical evergreen broad-leaved forest.

Ginkgo biloba, a deciduous tree distributed mainly along roadsides, was selected as the detection target. All training data and test data were collected from the three study sites, where there are many *Ginkgo biloba* trees.

2.2. Architecture. Figure 2 illustrates a general flowchart of the proposed architecture for detecting crowns and estimating morphological parameters automatically. A detailed explanation is given as follows.

Data are critical for the architecture, and the following prerequisites must be met: an orthophoto and a canopy

(a) (b) (c)

Figure 1: Study sites: (a) residential area,(b) western shoreline of Qingshanhu Lake, and (c) campus of ZAFU in Linan (top).

height model (CHM). The orthophoto, i.e., an image with a perfectly straight-down view of all objects, is needed to estimate crown size, whereas the CHM, which is a surface height distribution map, is essential for extracting tree height. Fortunately, we found that both an orthophoto and a CHM can be synthesized from a set of RGB images with a high overlap rate.

Deep learning, a general family of methods that use multilayered neural networks, has been proven effective at classifying, detecting, or segmenting objects in images. Both object detection neural networks and instance segmentation neural networks can be applied to this architecture (see Section 4.2). The deep learning system outlined in Figure 2 integrates all the functions needed, including data input, preprocessing, neural network construction, training, validation, analysis, and output. Before detection, the neural network should be trained based on training data; all trained parameters are stored in a model file. During detection, the system completes the following operations in order: preprocess the input orthophotos, build deep neural networks according to the model file, detect tree crowns in the neural network, and output the detection results. An object detection neural network, FPN-Faster-RCNN, is used to validate the architecture in this study.

The bounding box (bbox, Figure 3(a)) is the universal output form of object detection networks. (The output form of object segment networks is the bbox and mask; see Section 4.2.) We can obtain numbers and locations from the bbox statistics, which can also be used to estimate crown size and tree height.

Traditionally, crown size is represented by crown width, crown diameter, or crown area, which can be calculated by the bbox or mask. For example, crown width (a mean of two measurements made along north-south and east-west orientations; W. Lin et al. [34]; Minckler and Gingrich [35]; Vaz Monteiro et al. [36]) simply equals the average length of the sides of the bbox (see Equation (1)). (The top of all images in this study corresponds to north.)

$$\text{Crown width} = \text{average}\,(\text{length of sides of the bbox}). \quad (1)$$

The maximum pixel value of the corresponding area of a bbox (that is, the value of the brightest point in each bbox) in the CHM, a type of map named canopy height model, could be roughly estimated as the height of a tree (Figure 3(b)). It is easy to locate the position of the bbox in the CHM because the CHM and orthophoto originate from the same set of RGB images.

If the data and model are ready, the three tasks of detecting tree crowns, estimating crown widths, and estimating tree heights can be carried out continuously in one program without additional intervention.

2.3. Data Collection. The UAV used in this study was a DJI Inspire 2 (DJI Technology Co., Ltd., Shenzhen, China) and included four parts: the aircraft, remote control, camera, and power supply. All missions were flown at a height of 30-100 m above the launch site. The longitudinal overlap and side overlap were all set to 90%. The camera was set to orient vertically toward the surface. The total flight time for each individual flight was less than 15 min. All flights were conducted under light-wind or no-wind conditions.

Table 1 provides a summary of the data collected. A total of seven typical scenes, T1-T7, were chosen as test scenes (Figure 4), with T1 being from study site A (Figure 1), T2 being from study site B, and T3-T7 being from site C. Training data were also collected from sites B and C, therein excluding test scenes. To calculate crown size correctly, the images used for testing were orthophotos, which were synthesized from overlapping UAV images. However, the images used to create the training data included not only the orthophotos but also the original UAV images. Both types of images were used because the performance of a neural network typically improves as the quality, diversity, and amount of training data increase [37]. Although a large proportion of the tree crowns in the original drone images were likely to be taken from an oblique angle, the images were used to increase the data diversity and volume. Since there were few orthophotos available for training, most of the images used for training were original drone images.

The 7 test scenes are summarized in Table 2. The reference data on tree height were derived from field measurements obtained with the triangulation method using a laser rangefinder. The ground truth of the crown widths was obtained from manual measurements in the orthophotos instead of in the field because the resolution of the orthophotos was very high; thus, the data obtained from the orthophotos were expected to be more accurate than field measurements would be.

2.4. Data Processing. As shown in Figure 5, the point cloud, which is generated from overlapping images, comprises 3D point data and can be used to create orthophotos and DSMs directly. The digital terrain model (DTM), which represents the terrain surface, can be generated from point clouds by

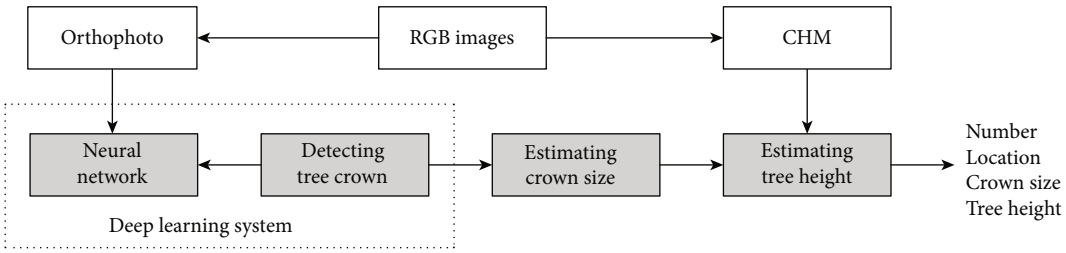
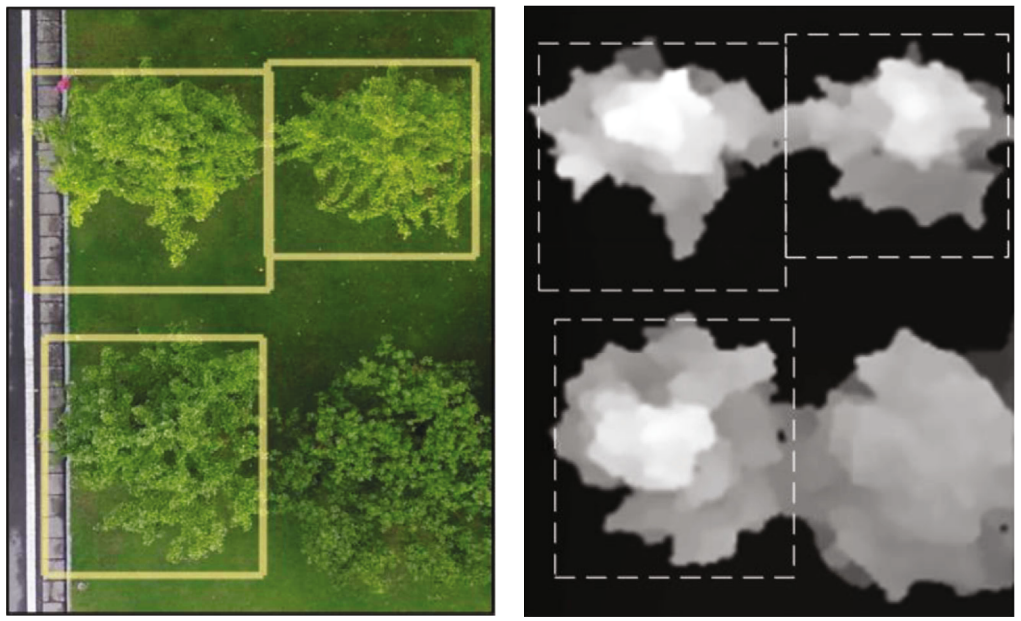


Figure 2: Flowchart of the proposed architecture: RGB: red-green-blue; CHM: canopy height model.



(a) (b)

Figure 3: Bounding boxes (bboxes), the output form of object detection networks: (a) bboxes in an orthophoto; (b) corresponding area in the CHM.

Table 1: Summary of data.

Data	Study sites	Acquisition time	Orthophotos	Original UAV images	Crowns
Training	B, C	2018.6/2019.6/2019.9	10	219	2593
Test	A, B, C	2019.6/2019.9	7	/	213

extracting the lowest points and interpolating them using the inverse distance weighting (IDW) method. Then, the CHM, which records tree height and other features, equals the difference between the DSM and DTM (see Equation (2)) [7]. If a point is the apex of a tree, we can obtain its elevation from the DSM, its ground elevation from the DTM, and tree height from the CHM.

$$\mathrm{CHM} = \mathrm{DSM} - \mathrm{DTM}. \quad (2)$$

The collected photos were processed with the 3D modeling software Agisoft PhotoScan Pro version 1.5.1 (Agisoft LLC, Russia). This software was chosen because it has proven effective in the production of mosaicked orthorectified imagery [38]. The data were processed with continuous operations, including photo alignment, alignment optimization, construction of dense point clouds, and orthophoto, DSM, and DTM construction. Finally, the CHM was created from ArcGIS operations.

2.5. FPN-Faster R-CNN. Faster region-based convolutional neural network (Faster R-CNN) is an object detection network based on convolutional neural networks developed by Ren et al. [39]. As shown in Figure 6, the input for Faster R-CNN is an image, and the output is bboxes around the objects identified by the Faster R-CNN program. The first module of Faster R-CNN is convnet, the output of which is a set of feature maps. The second module is the region proposal network (RPN), which generates a list of bboxes of likely positions of targets. More likely bboxes are stored in the region of interest (ROI) pool as candidate bboxes. The last module, classifier, consists of fully connected layers that

Figure 4: Detection results for ginkgo trees (yellow boxes for TP, red boxes for FP, and blue boxes for FN).

Table 2: Summary of test scenes.

Test scene	Acquisition	C1	C2	C3	C4	C5	C6	C7
T1	2019.9	14	9.07	11.00	6.56	5.21	6.52	3.72
T2	2019.9	20	8.02	11.00	5.43	3.84	4.93	2.44
T3	2019.6	42	9.36	12.40	5.78	3.62	5.85	1.26
T4	2019.6	40	9.32	13.90	5.46	3.76	6.56	1.32
T5	2019.6	22	9.69	12.50	4.91	4.29	5.32	3.42
T6	2019.6	52	10.96	14.21	6.60	4.24	6.10	2.88
T7	2019.9	23	11.29	13.82	9.10	5.64	7.37	4.25

C1: number of ginkgo trees; C2: mean tree height from field measurement; C3: maximum tree height from field measurement; C4: minimum height from field measurement; C5: mean width of measurement from orthophotos; C6: maximum width of measurement from orthophotos; C7: minimum width of measurement from orthophotos.

determine and output the optimal object categories and bboxes based on the loss function. Faster R-CNN has proven very efficient for detection, and its details can be found in the literature [39–41].

In Faster R-CNN, only the small-scale convnet feature layer is used for object detection. The layer is better at identifying simple objects than uneven distributions, different sizes, and overlapping tree crowns.

Many studies have found that small-sized feature layers are more conducive to extracting low-resolution, semantically strong features, whereas large-scale feature layers are effective at distinguishing high-resolution, semantically weak features. Feature pyramid network (FPN) is a novel structure that combines a small-scale feature layer and large-scale feature layers via a top-down pathway and lateral connections [42] (Figure 7). The top-down pathway creates large-scale feature layers by upsampling from the small-scale feature layer in higher pyramid levels. The newly created layers are then enhanced with feature layers from the bottom-up pathway via lateral connections. Each top-down feature layer can be seen as a collection of objects that can be extracted in lateral nets. This architecture can enhance the semantics of high-resolution objects, and it is very suitable for detecting uneven distributions, different sizes, and overlapping tree crowns.

FPN can be merged into other networks to improve their performance. Figure 7 illustrates the networks of FPN-Faster R-CNN that was adopted in this research to detect trees.

The training was based on a pretrained model file. Some of the hardware and software parameters for model training are shown in Table 3.

2.6. Accuracy Validation. The overall performance of tree identification and delineation was evaluated using the precision, recall, and F-score. Precision means the correct proportion of all detected objects, and recall refers to the correct proportion of all objects that should be detected. The F1-score, which is a harmonic mean of precision and recall, is used to refer to the overall accuracy [29, 43].

The overall accuracy F-score is defined by

$$\text{F-score} = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})}. \tag{3}$$

The precision and recall are defined as follows:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})}, \\ \text{Recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})}. \end{aligned} \tag{4}$$

TP, FP, FN, and IOU are defined as follows:

$$\begin{aligned} &\text{TP} = \text{number of detections with IOU} \geq 0.5, \\ &\text{FP} = \text{number of detections with IOU} \\ &\quad < 0.5 \text{ or detected more than once}, \\ &\text{FN} = \text{number of objects not detected}, \\ &\text{IOU} = \frac{(\text{detection result} \cap \text{ground truth})}{(\text{detection result} \cup \text{ground truth})}. \end{aligned} \tag{5}$$

The errors of the crown width and tree height estimation were evaluated by the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE), which are given by Equations (6), (7), and (8) as follows (where t_i is the true value and e_i is the estimate):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|t_i - e_i|, \tag{6}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|t_i - e_i|}{t_i}, \tag{7}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - e_i)^2}. \tag{8}$$

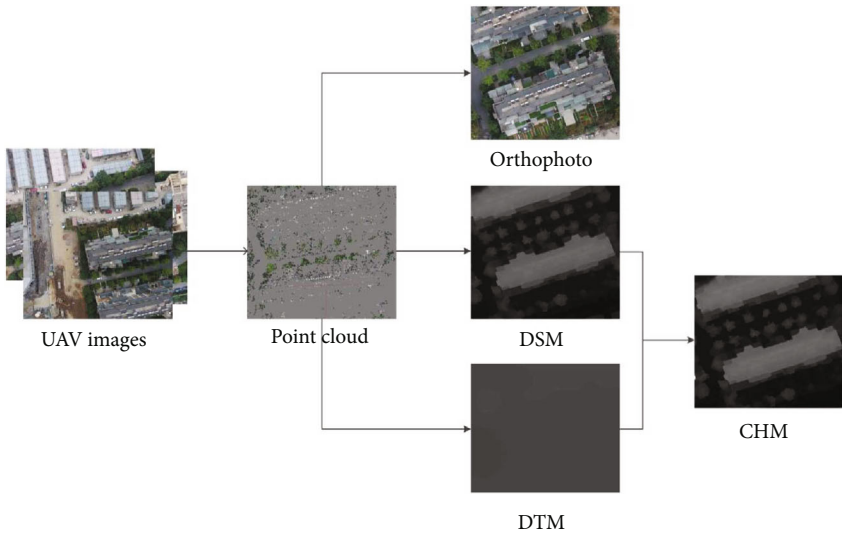


Figure 5: Data processing: DSM: digital surface model; DTM: digital terrain model; CHM: canopy height model. The DSM and DTM are both types of digital elevation model (DEM) and represent the elevation distribution of the region. They differ in that the pixel values in the DTM are the ground elevation, whereas those in the DSM are the elevation of the top surfaces of trees, buildings, and other aboveground features. Since the urban bare ground is very flat and the DTM is created by interpolation, the elevation values everywhere in the DTM are similar, and the resulting image just appears uniformly gray.

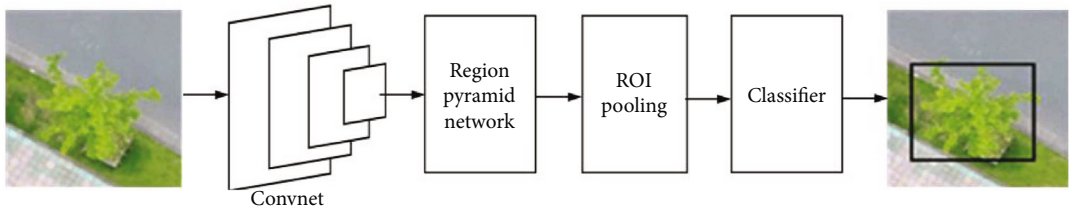


Figure 6: Faster R-CNN.

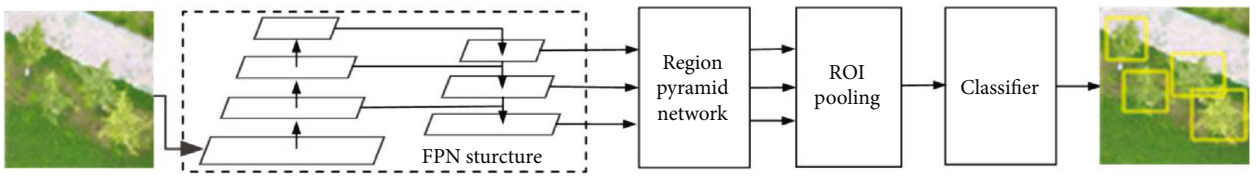


Figure 7: FPN-Faster R-CNN.

Table 3: Model training parameters.

	Parameter	Setting
Hardware	CPU	Intel Xeon E3-1225 V5 3.3 GHz
	GPU	Nvidia GeForce GTX 1080TI
	RAM	Hynix 32G
Software	Development language	Python
	Machine learning environment	TensorFlow
	Training iteration	18000

3. Results

3.1. Tree Crown Detection. In this study, 7 typical small scenes (Figure 4) were selected for testing because the use of only one large test area, as employed in previous studies, might be insufficient for representing urban landscapes. Test scene T1 represented a typical road, T2 represented a typical residential area, T3 and T4 represented areas containing tree crowns of different sizes, T5 and T6 represented areas with many overlapping canopies, and T7 represented an area where the color of the lawn background was similar to that of the crowns. Test scenes T3-T7, which were of the campus of Zhejiang A&F University and a botanical garden, were

Table 4: Detection results.

Test orthophotos	Total trees	Total detected	TP	FP	FN	Precision	Recall	*F*-score
T1	14	13	13	0	1	1	0.92	0.96
T2	20	19	19	0	1	1	0.95	0.97
T3	42	35	34	1	8	0.97	0.81	0.88
T4	40	28	28	0	12	1	0.70	0.82
T5	22	19	18	1	4	0.95	0.82	0.88
T6	52	50	44	6	8	0.88	0.85	0.86
T7	23	18	18	0	5	1	0.78	0.88
Total	213	182	174	8	39	0.96	0.82	0.88

TP = number of detections with IOU ≥ 0.5; FP = number of detections with IOU < 0.5 or detected more than once; FN = number of objects not detected; precision = TP/(TP + FP); recall = TP/(TP + FN); F-score = 2 ∗ precision ∗ recall/(precision + recall).

Table 5: Error causes.

Type	Total trees	Total detected	TP	FP	FN	Precision	Recall	*F*-score
Easy crowns	72	70	70	0	2	1	0.97	0.99
Hard crowns	141	112	104	8	37	0.93	0.74	0.78
Small crowns (width < 3 m)	28	15	15	0	13	1	0.54	0.70
Overlapping crowns	70	56	51	5	19	0.91	0.73	0.81
Lawn background	23	18	18	0	5	1	0.78	0.88

TP = number of detections with IOU ≥ 0.5; FP = number of detections with IOU < 0.5 or detected more than once; FN = number of objects not detected; precision = TP/(TP + FP); recall = TP/(TP + FN); F-score = 2 ∗ precision ∗ recall/(precision + recall).

challenging to detect. As shown in Figure 4, the results were acceptable.

Table 4 presents the detection results for each image. There were more than 400 tree crowns in the images. Comparing the detected objects with the 213 actual ginkgo trees, we found that 174 were correctly identified, with 39 false negatives; the recall was 0.82, the precision was 0.96, and the *F*-score was 0.88.

The impacts of natural factors such as crown size, canopy overlap, and background complexity on detection accuracy, which have been mentioned in previous studies, were the first issue to consider and are discussed here. For convenience, the two rows of ginkgo trees lining the roads and all the trees in T1 and T2 were classified as easy targets, and all other tree crowns were classified as hard targets. Table 5 shows the evaluation results for the easy targets and hard targets.

Most of the commission and omission errors were related to small crowns in T3 and T4 and overlapping crowns in T3, T4, T5, and T6. However, the omission errors in T7 were attributed to grass background and sunlight. In general, the results were acceptable because the overlap was so serious that we could only identify some crowns through field measurement.

Despite the problems mentioned above, our method showed very good performance: (1) regardless of the scene, the accuracy in easy target detection was very high, indicating the stability of our method. (2) Although there were many other species of trees (more than 200) in the scenes, our method incorrectly classified other species of trees as ginkgo trees only 3 times, demonstrating excellent classification performance. (3) Most artificial features, such as buildings and roads, did not interfere with the recognition of ginkgo trees. There was only one omission error due to artificial features, which was in the lower left corner in T2 and related to obstruction by a street lamp. (4) The uneven distribution of tree crowns excluding overlapping had a minimal effect on the test results.

3.2. Crown Width. Only correctly detected ginkgo trees are discussed in this section. As mentioned before, instead of field-measured crown width, manual measurements from the orthophotos were used as reference data because of the high resolution of the orthophotos. For crown width estimation, the MAE was 0.37 m, the MAPE was 8.71%, and the RMSE was 0.495 m. The largest error percentage rate of the crown width estimation was 37.2%.

Figure 8(a) shows the relationship between manually measured crown width and automatically estimated crown width. Figure 8(b) shows the percentage error (absolute (estimated – manually measured)/manually measured ∗ 100%) distribution. The results showed good agreement between the ground truth and the estimates. Some points far from the line $y = x$ correspond to overlapping crowns.

3.3. Tree Height. As in the previous section, only correctly detected ginkgo trees are discussed in this section. The largest absolute percentage error of tree height obtained by automatic detection was 67.8%, the MAE was 0.68 m, the MAPE was 7.33%, and the RMSE was 0.987 m.

Figure 8(c) presents the relationship between field measurements of tree height and automatically estimated tree height. Figure 8(d) shows the percentage error (absolute (estimated – field measured)/field measured ∗ 100

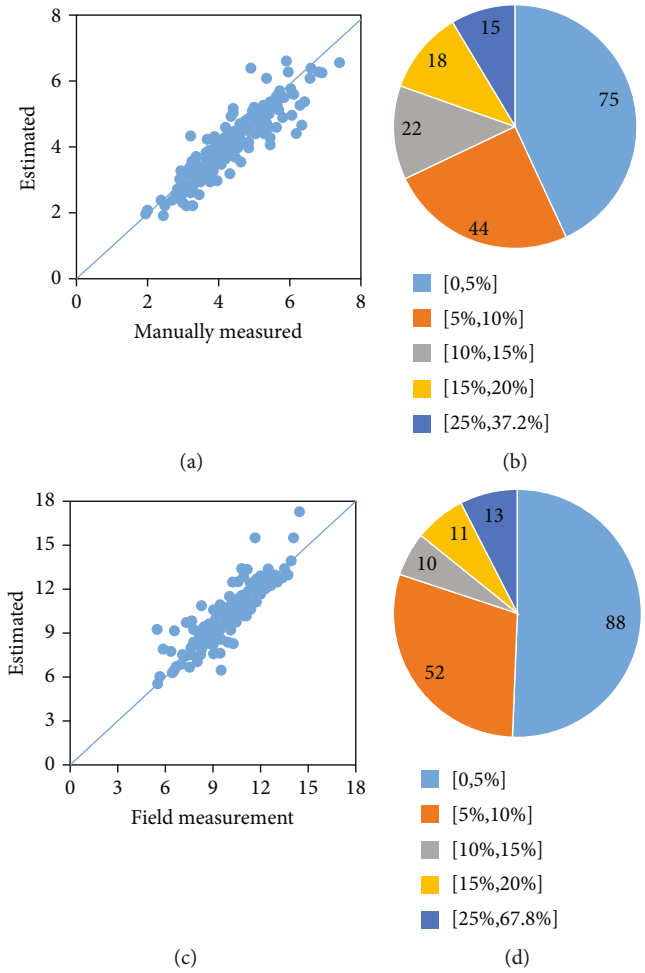


Figure 8: (a) Estimated versus manually measured crown width (both in meters); (b) percentage error distribution of crown width estimation; (c) estimated versus field-measured tree height (both in meters); (d) percentage error distribution of tree height estimation.

%) distribution. The results showed good agreement between the ground truth and the estimates. The points far from the line $y = x$ might have been caused by unpredictable interference factors.

3.4. IOU. In this study, IOU was used as an assessment criterion for determining whether a crown was detected. This criterion has been used in a few recent studies [44, 45]. The IOU threshold ultimately selected for use in this paper was 0.5, which is the general standard in deep learning studies. At higher thresholds, the detection accuracy decreased, and at a lower threshold, the detection accuracy increased (Table 6). However, the detection accuracy improved minimally when we lowered the threshold to 0, which means that very few (only 4) bbox IOUs were within [0, 0.5], illustrating the great performance of FPN-Faster R-CNN.

The IOU threshold also had a strong impact on the error statistics of crown width because crown width was calculated from the bbox. At higher thresholds, the MAE and MAPE of crown width decreased.

Table 6: Impacts of IOU.

Criterion	Detection precision	Detection recall	Detection *F*-score	MAPE of crown width	MAPE of tree height
IOU > 0	0.97	0.83	0.90	9.19%	7.50%
IOU > 0.5	0.96	0.82	0.88	8.71%	7.33%
IOU > 0.6	0.84	0.71	0.77	6.90%	7.36%
IOU > 0.7	0.67	0.57	0.61	5.97%	7.29%

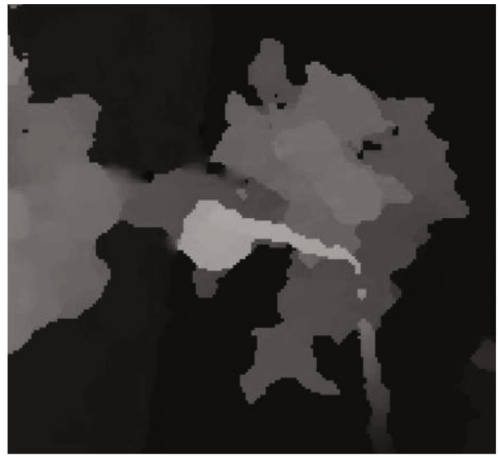


Figure 9: A street lamp above a ginkgo tree.

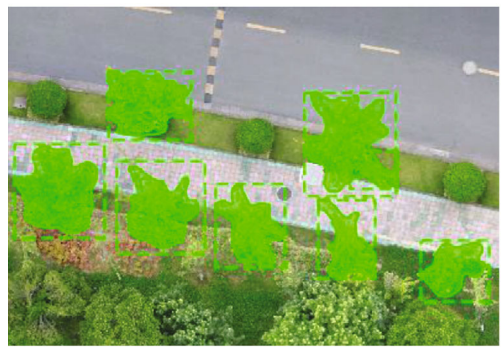


Figure 10: A result of Mask-RCNN.

We believe that the accuracy of the tree height estimation was affected mainly by the following aspects: (1) the error from the field measurement and the CHM map, (2) whether the highest point of the tree was within the bbox, and (3) interference from artificial objects. If a bbox had a large deviation (IOU < 0.5), the highest point of the tree may not be inside it, resulting in an incorrect estimation of tree height. When IOU > 0.5, the search for the highest point of a canopy usually yielded correct results and the error was mainly derived from the field measurements and CHM maps. In addition, there are countless complex scenes and unpredictable interfering factors in cities. For example, as shown in Figure 9, a large deviation in tree height estimation occurred due to a street lamp.

Table 7: Studies on individual tree detection and parameter estimation.

Study	1 [30]	2 [31]	3 [29]
Sensor	2 laser scanners, two CCD cameras	ADS40 airborne digital sensor	Submeter optical sensor
Carrying platform	Van	Plane	WorldView-2 satellite
Detection method	Voxel-based marked neighborhood searching	Classification of refined superpixels by a naive Bayes classifier	Marked point processes based on a geometrical- optical model
Study area	Small	Large	Large
Target	Individual street trees	Individual urban tree crowns	Individual urban tree crowns
Species recognition	No	No	No
Crown size	Y	Y	Y
Height	Y	Y	N
DBH	Y	N	N

4. Discussion

This paper presents a novel method to conveniently and efficiently map the individual number, locations, crown sizes, and tree heights of *Ginkgo biloba* trees.

4.1. Data Sources. Some common data sources were available for our study; for example, orthophotos can be synthesized from RGB, multispectral, or high spectral images, and LIDAR data or high overlapping RGB images are often used to create a CHM. Considering the powerful detection capabilities of deep learning, in this research, we chose a convenient, low-cost data solution in which orthophotos and CHMs are all created from highly overlapping RGB images. This approach has not often been adopted in previous studies due to the poor spectral information of such images.

However, the quality of several orthophotos was not ideal because there were some distorted crowns, especially in the dense area, due to the orthophoto synthesis process. Moreover, a few errors in the test were attributed to synthesis quality (e.g., the omission errors in the lower right corner of T6).

In addition, our program cannot yet handle large images. Therefore, if the spatial extent of an orthophoto is very large, it should be divided into small images before detection.

4.2. Flexibility of the Architecture. Any new object detection network or instance segmentation network can be applied to this architecture, such as FPN-Faster R-CNN, because object detection neural networks (such as Faster R-CNN and YOLO) output a bbox around each target, and instance segmentation networks (such as Mask R-CNN and BlendMask) all output the masks of the canopies as well as bboxes (Figure 10). We selected FPN-Faster R-CNN for our case study because of its ease of discussion, although it would have been possible for us to obtain another crown size indictor (crown area) from an instance segmentation network such as Mask-RCNN.

Therefore, the architecture has great potential for performance improvement because of this flexibility. In recent years, an increasing number of studies have focused on deep learning, and new high-performance neural networks are being continuously developed. More accurate results would be achieved if high-performance networks were adopted in this architecture.

4.3. Comparison. There have been a few schemes focused on building architectures to realize individual urban tree detection and parameter estimation simultaneously. The methods and sensors of some representative studies are listed in Table 7. In one study (study 1 in Table 7), a mobile solution with many sensors was presented; the design made it costly and limited its observable range. However, its detection of street trees (such as easy targets in Table 5) was over 98%, and the RMSE values of its estimation of tree height, DBH, and crown diameter were 0.15 m, 0.01 m, and 0.13 m, which shows its good capability in parameter estimation. While the observable range of study 2 is comparatively large, so are its costs. Additionally, there were not enough accuracy estimates suitable for comparison with our study. Given its huge coverage and the low image resolution of WorldView-2, the method presented in study 3 achieved acceptable evaluation results, with detection and delineation accuracies of 0.87 and 0.63, respectively. Compared with these schemes, our method has several benefits that cannot be ignored, such as convenience, low cost, capability of species recognition, and tremendous potential for performance improvements based on data growth and methodology evolution. Nonetheless, how to estimate DBH from the air remains an unsolved problem.

In addition to the above schemes, another scheme is available that can realize individual tree detection and parameter estimation automatically. In this scheme, both tree height estimation and tree crown detection are conducted based on CHM by using the method of local maxima. Guerra-Hernández et al. [46] validated the scheme in an area of umbrella pine afforestation, and all the trees of the plots were correctly detected. The RMSE values for the predicted heights and crown widths were 0.45 m and 0.63 m, respectively. Nonetheless, successful validation with individual urban trees has been absent, possibly due to the presence of urban infrastructure, which can result in ambiguity when height ranges are extracted to estimate tree height [6].

In addition, some studies have only focused on developing a new method to detect individual urban trees. Lin et al.

[28] developed a nondeep learning method based on UAV oblique images, and the commission and omission errors were less than 32% and 26%. Xie et al. [27] presented a two-phase deep learning method to detect urban trees based on normal height model (NHM) images, which achieved F-scores between 85% and 90%. Torres et al. [47] evaluated five deep fully convolutional networks (FCNs) for the semantic segmentation of a single tree species: SegNet, U-Net, FC-DenseNet, and two DeepLabv3+ variants. The experimental analysis revealed the effectiveness of each design, with F-scores ranging from 87.0% to 96.1%. Although all these results are based on different conditions, we believe that the accuracy of FPN-Faster RCNN adopted in this study is higher than that of the above nondeep learning method and is roughly at the same level as other deep learning networks.

4.4. Application Scenes. In general, our architecture is suitable for urban environments for 2 reasons: (1) complex urban features have little influence on deep learning-based tree canopy detection and (2) the flat terrain in urban areas is conducive to obtaining accurate tree height values. However, small patches of dense woodland in cities are a challenge for our framework, which can be attributed to its use of RGB images, which provide only color and texture information. Differences in texture and color between overlapping crowns are not obvious in many cases, which is a difficult problem for deep learning.

We believe that our method can be easily extended to plantations or sparse natural forests, grasslands, pastures, and other areas where canopy overlap is not extensive. However, if the terrain is complex, the estimation of height will be slightly affected; Guerra-Hernández et al. [48] estimated tree height based on CHM data in an area with complex terrain and obtained an RMSE of 2.84 m, which is not ideal.

We believe that in dense natural forest, severe crown overlap will affect not only the accuracy of crown detection but also the estimation accuracy of tree height and crown size. A feasible solution is to not only use CHM data for tree height estimation but also superimpose them with RGB images to train the deep learning networks. This approach is believed to be effective, although some experiments with this method show that there remains room for improvement [49].

5. Conclusions

In this paper, a novel method for automatic tree crown detection and parameter estimation using deep learning technology is proposed, and FPN-Faster R-CNN is used in a deep learning example to verify the architecture. The method realizes automatic tree crown detection and morphology parameter estimation in some complex urban scenes and is convenient and low cost. The quality of the orthophotos affected the canopy detection results. In general, deep learning is a very promising method that warrants further research, and the accuracy of the information collected by the architecture will increase as neural networks evolve.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant number U1809208); Joint Funds of the Natural Science Foundation of Zhejiang Province, China (grant number LQY18C160002); Natural Science Foundation of Zhejiang Province, China (grant number LQ20F020005); and Zhejiang Science and Technology Key R&D Program Funded Project (grant number 2018C02013).

References

[1] G. Kuchelmeister and S. Braatz, "Urban forestry revisited," *Unasylva*, vol. 44, 1993.

[2] E. G. McPherson, D. J. Nowak, and R. A. Rowntree, *Chicago's Urban Forest Ecosystem: Results of the Chicago Urban Forest Climate Project*, U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, 1994.

[3] S. Roy, J. Byrne, and C. Pickering, "A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones," *Urban Forestry & Urban Greening*, vol. 11, no. 4, pp. 351–363, 2012.

[4] E. Zimmermann, L. Bracalenti, R. Piacentini, and L. Inostroza, "Urban flood risk reduction by increasing green areas for adaptation to climate change," vol. 161, pp. 2241–2246, 2016.

[5] M. A. Goddard, A. J. Dougill, and T. G. Benton, "Scaling up from gardens: biodiversity conservation in urban environments," *Trends in Ecology & Evolution*, vol. 25, no. 2, pp. 90–98, 2010.

[6] L. Liu, N. C. Coops, N. W. Aven, and Y. Pang, "Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data," *Remote Sensing of Environment*, vol. 200, pp. 170–182, 2017.

[7] K. Iizuka, T. Yonehara, M. Itoh, and Y. Kosugi, "Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese cypress (Chamaecyparis obtusa) forest," *Remote Sensing*, vol. 10, no. 2, p. 13, 2018.

[8] J. Hyyppä and M. Inkinen, "Detecting and estimating attributes for single trees using laser scanner," *Photogramm J Finl*, vol. 16, pp. 27–42, 1999.

[9] M. Wulder, K. O. Niemann, and D. G. Goodenough, "Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery," *Remote Sensing of Environment*, vol. 73, no. 1, pp. 103–114, 2000.

[10] T. Brandtberg, T. A. Warner, R. E. Landenberger, and J. B. McGraw, "Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America," *Remote Sensing of Environment*, vol. 85, no. 3, pp. 290–303, 2003.

[11] J. Pitkänen, "Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima

methods," *Canadian Journal of Forest Research*, vol. 31, no. 5, pp. 832–844, 2001.

[12] P. Bunting and R. Lucas, "The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data," *Remote Sensing of Environment*, vol. 101, no. 2, pp. 230–248, 2006.

[13] L. Wang, P. Gong, and G. S. Biging, "Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 3, pp. 351–357, 2004.

[14] F. A. Gougeon and D. G. Leckie, "The individual tree crown approach applied to Ikonos images of a coniferous plantation area," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 11, pp. 1287–1297, 2006.

[15] Z. Zhen, L. J. Quackenbush, and L. Zhang, "Trends in automatic individual tree crown detection and delineation-evolution of LiDAR data," *Remote Sensing*, vol. 8, no. 4, 2016.

[16] Y. Ke and L. J. Quackenbush, "A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing," *International Journal of Remote Sensing*, vol. 32, no. 17, pp. 4725–4747, 2011.

[17] M. Dalponte, H. O. Ørka, L. T. Ene, T. Gobakken, and E. Næsset, "Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data," *Remote Sensing of Environment*, vol. 140, pp. 306–317, 2014.

[18] J. Maschler, C. Atzberger, M. Immitzer, J. Maschler, C. Atzberger, and M. Immitzer, "Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data," *Remote Sensing*, vol. 10, no. 8, p. 1218, 2018.

[19] A. C. Birdal, U. Avdan, and T. Türk, "Estimating tree heights with images from an unmanned aerial vehicle," *Hazards Risk*, vol. 8, no. 2, pp. 1144–1156, 2017.

[20] S. Koukoulas and G. A. Blackburn, "Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data," *International Journal of Remote Sensing*, vol. 26, no. 3, pp. 431–455, 2005.

[21] Y. S. Lim, P. H. La, J. S. Park, M. H. Lee, M. W. Pyeon, and J. I. Kim, "Calculation of tree height and canopy crown from drone images using segmentation," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 33, no. 6, pp. 605–614, 2015.

[22] F. Morsdorf, E. Meier, B. Kötz, K. I. Itten, M. Dobbertin, and B. Allgöwer, "LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management," *Remote Sensing of Environment*, vol. 92, no. 3, pp. 353–362, 2004.

[23] R. A. Díaz-Varela, R. de la Rosa, L. León, and P. J. Zarco-Tejada, "High-resolution airborne UAV imagery to assess olive tree crown parameters using 3D photo reconstruction: application in breeding trials," *Remote Sensing*, vol. 7, no. 4, pp. 4213–4232, 2015.

[24] P. Shin, T. Sankey, M. Moore, and A. Thode, "Evaluating unmanned aerial vehicle images for estimating forest canopy fuels in a ponderosa pine stand," *Remote Sensing*, vol. 10, no. 8, p. 1266, 2018.

[25] L. Wallace, A. Lucieer, Z. Malenovský et al., "Assessment of forest structure using two UAV techniques: a comparison of airborne laser scanning and structure from motion (SfM) point clouds," *Forests*, vol. 7, no. 12, p. 62, 2016.

[26] G. Mozgeris, V. Juodkienė, D. Jonikavičius, L. Straigytė, S. Gadal, and W. Ouerghemmi, "Ultra-light aircraft-based hyperspectral and colour-infrared imaging to identify deciduous tree species in an urban environment," *Remote Sensing*, vol. 10, no. 10, p. 1668, 2018.

[27] Y. Xie, H. Bao, S. Shekhar, and J. Knight, "A TIMBER framework for mining urban tree inventories using remote sensing datasets," in *in: 2018 IEEE international conference on data mining*, pp. 1344–1349, Singapore, 2018.

[28] Y. Lin, M. Jiang, Y. Yao, L. Zhang, and J. Lin, "Use of UAV oblique imaging for the detection of individual trees in residential environments," *Urban Forestry & Urban Greening*, vol. 14, no. 2, pp. 404–412, 2015.

[29] M. F. Gomes, P. Maillard, and H. Deng, "Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model," *Remote Sensing of Environment*, vol. 211, pp. 184–195, 2018.

[30] B. Wu, B. Yu, W. Yue et al., "A voxel-based method for automated identification and morphological parameters estimation of individual street trees from mobile laser scanning data," *Remote Sensing*, vol. 5, no. 2, pp. 584–611, 2013.

[31] J. Jiao and Z. Deng, "Individual building rooftop and tree crown segmentation from high-resolution urban aerial optical images," *Journal of Sensors*, vol. 2016, 13 pages, 2016.

[32] G. Morales, G. Kemper, G. Sevillano, D. Arteaga, I. Ortega, and J. Telles, "Automatic Segmentation of Mauritia flexuosa in Unmanned Aerial Vehicle (UAV) Imagery Using Deep Learning," *Forests*, vol. 9, no. 12, p. 736, 2018.

[33] Y. Ampatzidis and V. Partel, "UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence," *Remote Sensing*, vol. 11, no. 4, p. 410, 2019.

[34] W. Lin, Y. Meng, Z. Qiu, S. Zhang, and J. Wu, "Measurement and calculation of crown projection area and crown volume of individual trees based on 3D laser-scanned point-cloud data," *International Journal of Remote Sensing*, vol. 38, no. 4, pp. 1083–1100, 2017.

[35] L. S. Minckler and S. F. Gingrich, *Relation of Crown Width to Tree Diameter in Some Upland Hardwood Stands of Southern Illinois*, 1970.

[36] M. Vaz Monteiro, K. J. Doick, and P. Handley, "Allometric relationships for urban trees in Great Britain," *Urban Forestry & Urban Greening*, vol. 19, pp. 223–236, 2016.

[37] H. Salehinejad, S. Valaee, T. Dowdell, and J. Barfett, "Image Augmentation Using Radial Transform for Training Deep Neural Networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3016–3020, Calgary, AB, Canada, 2018.

[38] J. Dempewolf, J. Nagol, S. Hein et al., "Measurement of within-season tree height growth in a mixed forest stand using UAV imagery," *Forests*, vol. 8, no. 7, p. 231, 2017.

[39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[40] R. Girshick, "Fast R-CNN - IEEE conference publication," in *in: 2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2016.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[42] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944, 2017a.

[43] G. Goldbergs, S. Maier, S. Levick, and A. Edwards, "Efficiency of individual tree detection approaches based on light-weight and low-cost UAS imagery in Australian savannas," *Remote Sensing*, vol. 10, no. 2, p. 161, 2018.

[44] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks," *Remote Sensing*, vol. 11, no. 11, p. 1309, 2019.

[45] T. Zhao, Y. Yang, H. Niu, Y. Chen, and D. Wang, "Comparing U-Net convolutional network with mask R-CNN in the performances of pomegranate tree canopy segmentation," in *Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VII*, 2018.

[46] J. Guerra-Hernández, E. González-Ferreiro, A. Sarmento et al., "Using high resolution UAV imagery to estimate tree variables in Pinus pinea plantation in Portugal," *For. Syst.*, vol. 25, no. 2, 2016.

[47] D. L. Torres, R. Q. Feitosa, P. N. Happ et al., "Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery," *Sensors (Switzerland)*, vol. 20, no. 2, 2020.

[48] J. Guerra-Hernández, D. N. Cosenza, L. C. E. Rodriguez et al., "Comparison of ALS- and UAV(SfM)-derived high-density point clouds for individual tree detection in Eucalyptus plantations," *International Journal of Remote Sensing*, vol. 39, no. 15-16, pp. 5211–5235, 2018.

[49] A. I. Pleşoianu, M. S. Stupariu, I. Şandric, I. Pătru-Stupariu, and L. Drăguţ, "Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model," *Remote Sensing*, vol. 12, no. 15, p. 2426, 2020.