future internet

MDPI

Article

The Combined Use of UAV-Based RGB and DEM Images for the Detection and Delineation of Orange Tree Crowns with Mask R-CNN: An Approach of Labeling and Unified Framework

Felipe Lucena [1,*], Fabio Marcelo Breunig [2] and Hermann Kux [1]

[1] Divisão de Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais (INPE), Av. dos Astronautas, 1758—Jardim da Granja, São José dos Campos CEP 12227-010, SP, Brazil
[2] Departamento de Engenharia Florestal, Universidade Federal de Santa Maria (UFSM)—Campus Frederico Westphalen, linha Sete de Setembro s/n, UFSM, Frederico Westphalen CEP 98400-000, RS, Brazil
* Correspondence: felipesa01@gmail.com

Abstract: In this study, we used images obtained by Unmanned Aerial Vehicles (UAV) and an instance segmentation model based on deep learning (Mask R-CNN) to evaluate the ability to detect and delineate canopies in high density orange plantations. The main objective of the work was to evaluate the improvement acquired by the segmentation model when integrating the Canopy Height Model (CHM) as a fourth band to the images. Two models were evaluated, one with RGB images and the other with RGB + CHM images, and the results indicated that the model with combined images presents better results (overall accuracy from 90.42% to 97.01%). In addition to the comparison, this work suggests a more efficient ground truth mapping method and proposes a methodology for mosaicking the results by Mask R-CNN on remotely sensed images.

Keywords: precision agriculture; instance segmentation; tree detection; tree delineation; UAV-based images; Mask R-CNN

Citation: Lucena, F.; Breunig, F.M.; Kux, H. The Combined Use of UAV-Based RGB and DEM Images for the Detection and Delineation of Orange Tree Crowns with Mask R-CNN: An Approach of Labeling and Unified Framework. *Future Internet* **2022**, *14*, 275. https://doi.org/10.3390/fi14100275

Academic Editors: Michael Sfakiotakis, Spyros Panagiotakis and Ioannis N. Daliakopoulos

Received: 26 July 2022
Accepted: 8 September 2022
Published: 27 September 2022

1. Introduction

Appropriate management techniques are an important help to remediate agricultural challenges regarding productivity, environmental impact, food security, and sustainability [1]. Due to the heterogeneity and complexity of agricultural environments, it is necessary to consider them, with regard to monitoring, measurement, and continuous analysis of the physical aspects and phenomena involved [2]. For agricultural fields cultivated with woody species, individual tree canopy identification is important for plant counts, its growth assessment, and yield estimation. On the other hand, tree detection and mapping are costly activities that consume a lot of time and effort when performed by traditional field techniques [3,4]. In view of this, practical approaches for the efficient identification and mapping of such trees are necessary to improve agricultural management.

Remote sensing data have been widely used for agricultural management as well as for plant analysis. The data of interest for such an evaluation is collected individually for each plant in the field. In particular, Unmanned Aerial Vehicles (UAVs) are the most frequently used remote sensing platforms for this activity because they provide images with refined spatial resolution, allowing for on-demand imaging, enabling timely processing and acquisition of information, and they are cost-effective compared to manned aerial or equivalent orbital imaging [5]. From tobacco plants with a focus on agricultural management using RGB images acquired by UAVs, Ref. [6] evaluated the ability of automatic detection. RGB UAV images were used by Ref. [7] to identify individual rice plants in agricultural fields to estimate productivity. Multispectral UAV images were used by Ref. [8] to derive spectral attributes (i.e., vegetation indices) for the identification of the greening disease in orange plantations through the recognition of individual canopies before its classification.

The imaging equipment commonly used in UAVs are RGB digital cameras, multi and hyperspectral sensors and, less commonly, thermal and light detection and ranging (LiDAR) systems [5]. The use of RGB cameras in UAV is considerably more accessible than later methods. Using photogrammetry techniques and algorithms, such as Structure from Motion (SfM) and Multi-View Stereo (MVS), allows for the acquisition of data beyond the imaged scenes, such as dense three-dimensional point clouds, Digital Elevation Models (DEM), and ortho-rectified image mosaics [9].

The integration of images with these three-dimensional data allows for robust vegetation monitoring, not only due to the ability to identify individual plants but also because of the possibility to estimate canopy morphology parameters, such as height, diameter, perimeter, and volume [10–12]. The canopy detection techniques from remote sensing images apply different concepts and its suitability depends on the type of canopy studied and the characteristics of the evaluated areas. The classic methods for this task are the Local Maxima (LM) algorithm, Marker-Controlled Watershed Segmentation (MCWS), template matching, region-growing, and edge detection [13]. However, the major limitation for the application of these methods is its requirement to manually configure specific parameters for each type of target of interest or image during identification [14,15] with analyst's specific knowledge and so it reduces the feasibility of developing automated process frameworks.

In addition to classical methods, Deep Learning (DL) models have been widely used in recent years due to its ability to deal with various computer vision problems [16], especially with Convolutional Neural Network (CNN) architecture. One of the advances made in the field of CNNs was the development of Region-based CNNs (R-CNN) [17], which is architecture able to perform image instance segmentation—the individual identification of objects belonging to the same semantic class (i.e., the discrimination of individual treetops within the vegetation class).

The latest advance of R-CNNs is the Mask R-CNN architecture [18], which is capable of accomplishing instantiated identification and delimitation of the contour from the object of interest at the pixel level. Mask R-CNN has been widely used in remote sensing of vegetation and the results obtained indicate its great potential for the detection and design of targets. A comparison of the performance of classic models with Mask R-CNN for the identification of China fir was made by Ref. [13], and it was concluded that Mask R-CNN presents a higher performance, reaching a F1-score up to 0.95. The parameter setting of the model for the use of training samples with different levels of refinement was investigated by Ref. [19], reaching a F1-score of 0.90 and 0.97 for potato and lettuce plantations, respectively.

Even with the application of DL, using region-based or other CNN models, the detection of tree canopies still requires further investigation, such as the use of a single model for areas with heterogeneous spatial characteristics or in cultures with high planting densities, where the contiguity and overlap between two neighboring crowns occurs naturally and makes individual identification a really difficult task [20,21]. On the other hand, studies of instance segmentation applied to natural scenes have shown that the performance of models is superior when combining scene depth information with RGB images [22]. Therefore, the structural/morphological information from data based on photogrammetry, linked to the crown images, can be a relevant factor for the identification of each grouped crowns.

To the best of our knowledge, few studies evaluated the use of three-dimensional information from RGB images as a proxy for individualization of contiguous treetops. the ability to identify individual crowns of chestnut trees from DEM derived products based on SfM was evaluated by Ref. [10]. The process suggested by the authors is based mainly based on the use of morphological filters since the chestnut crowns, despite touching each other, clearly maintain its circular path, as seen from nadir. Combinations of different RGB and SfM-based features for crown identification and Chinese fir height calculation were analyzed by Ref. [12]; however, similarly, crowns do not contemplate a high level of density. Furthermore, some cultures have different characteristics of the mentioned species and the

crowns are located extremely dense, such as orange trees and grapes. a tool based on a dense cloud of points and products derived from DEM for surveying viticulture biomass was developed by Ref. [11]. The authors did not focus on the individual identification of each plant to calculate the biomass but on the identification of failures due to the decrease in the canopy density.

An additional major challenge for the use of DL is the need of a significant amount and variety of training samples for network learning, which requires a lot of manual labeling work. Some studies are based on this condition to simplify the detection of trees, identified by an enclosing rectangle, without delimiting the contour of each crown [20,23] or even considering the punctual representation, without the two-dimensional delimitation of the plant [24–26]. Nevertheless, the identification of trees by a canopy delimiting mask offers a range of possibilities for the analysis due to the discrimination of the canopy area in in relation to its surroundings (i.e., soil and shade). It allows for the application of approaches aimed at individual location, tree counts, and the extraction of morphological information, as mentioned, but above all the use of the exclusive spectral response from the plant canopy, as in the identification of phyto-pathologies [27].

Another challenge for using models based on DL is the adequacy of the images to the architecture of the models, as these are usually built for natural images of fixed proportions and especially smaller sizes than orbital or UAV images. Therefore, the use of remote sensing images in these models requires the subdivision of the original image into several square patches that require post-treatment to eliminate the mosaic effect [6,13], which makes it difficult to develop automated detection and counting frameworks. To the best of our knowledge, few studies have addressed any methodology for aggregating the final result in a single scene. Through a modification of the Non-Maximum Suppression (NMS) method, Ref. [28] proposed an approach to study with irrigated pivots, which differ from trees in agricultural plots as they have extremely regular shapes and no overlap. To integrate the results of identification of forest tree canopies, Ref. [29] adopted their own method. Without much detail about the procedure, the authors suggest the union of any two crowns that overlap and are located at the ends of the clippings (patches), disregarding the possibility of real overlap between the trees.

In the existing literature, few studies have reported the evaluation of instance segmentation in remote sensing images using only RGB imagery and photogrammetry DEM-based data for detection and delineation of dense treetops (touching/overlapping). The analysis carried out in this study aims (I) to contribute to the identification and counting of plants and the delimitation of their canopies, as well as (II) to increase the feasibility of the annotation process in studies involving canopies of contiguous trees in agricultural orchards, and also (III) to propose a method for automating the segmentation process of canopies with different planting densities for large scenes by mosaicking the results.

2. Materials and Methods

2.1. Study Area

The study area (Figure 1) is composed of three fractions of plots located in São Paulo State, Brazil, referred as plots A (47.1285° W, 22.0543° S), B (47.0491° W, 22.4555° S) and C (47.0003° W, 22.4508° S). These plots cover 5.20, 4.33, 7.38 hectares, respectively, totaling 16.9 ha of orange plantations of Hamlin, Baianinha, Valencia, Pêra, and Natal varieties. Altogether, the segments of plots cover 9064 trees and have different spatial characteristics, such as planting age, crown height and diameter, spacing between trees and rows, as well as different soil coverage.

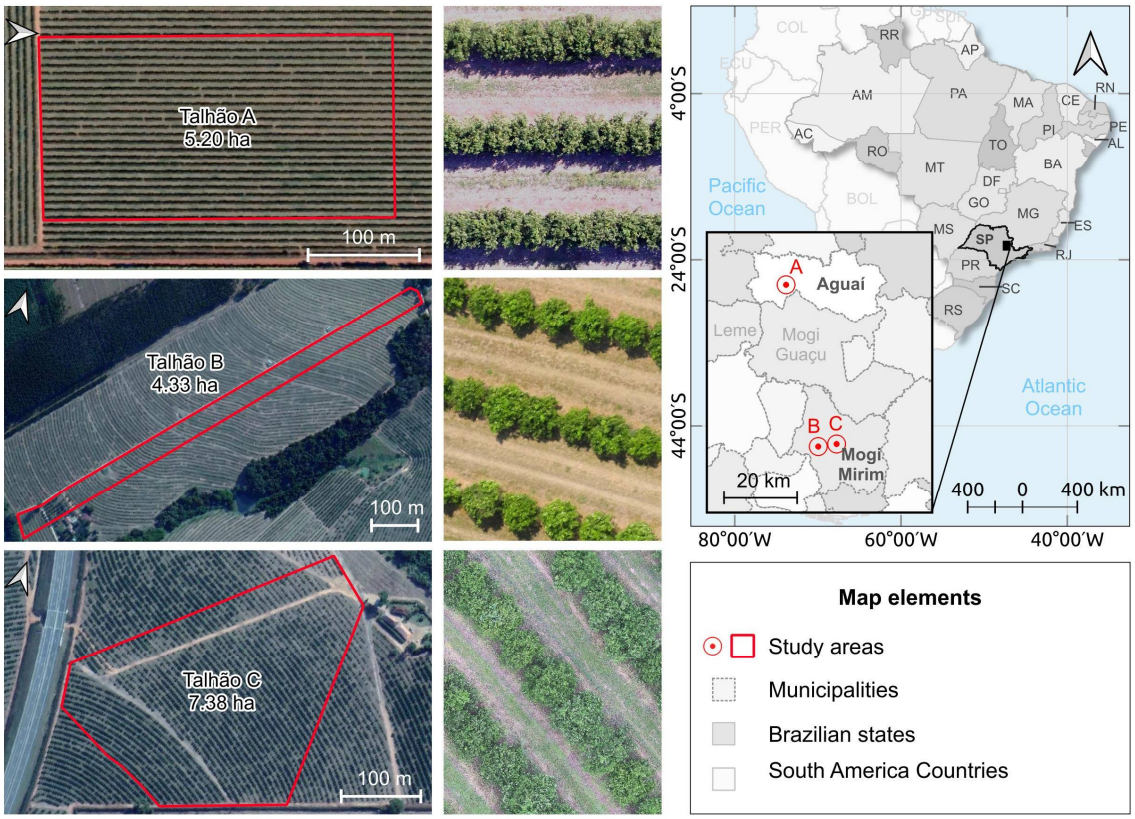

Figure 1. Location map of the study areas and details of the adopted plots.

2.2. Image Acquisition and Processing

The RGB images were acquired in three plots at different dates and with different equipment. Plots A and C were imaged with a multi-rotor UAV, the DJI Phantom 3 (DJI, Shenzhen, China) with an RGB digital camera—PowerShot S100 (zoom lens 5.2 mm; 12.1 Megapixel CMOS sensor; 4000 × 3000 resolution). Plot B was imaged with a fixed-wing UAV, senseFly's eBee (senseFly SA, Lausanne, Switzerland) integrated with the senseFly Duet T camera, which has an RGB sensorSensefly S.O.D.A. (zoom lens 35 mm; 5472 × 3648 resolution) and a thermal sensor that was not considered in this study. Table 1 summarizes information related to image acquisition, including equipment and flight configuration.

Table 1. Characteristics of the imaging flights of each studied plot.

Plot	Date	UAV	Fly Height (m)	GSD (cm)	Overlap (%) (Front/Side)
A	5 December 2019	DJI Phantom 3	50	2.2	80/70
B	27 September 2021	SenseFly's eBee	76	7.1	80/75
C	20 January 2020	DJI Phantom 3	50	2.5	80/80

Pix4Dmapper Pro software (Pix4D SA, Lausanne, Switzerland) was used for the photogrammetric processing of the images acquired by the UAVs, using sequentially the following techniques: alignment optimization, construction of dense mesh of points, classification of ground points, elaboration of DEM-based data (such as the Digital Surface Model (DSM) and Digital Terrain Model (DTM)) and ortho-mosaics. In addition to the products generated by the software, the Canopy Height Model (CHM) of each plot was also computed. The CHM is the result of the subtraction of the DSM by the DTM, generating an elevation model that considers only the heights above the ground and eliminates the terrain slope. The canopy height information is important to define the separation between two

dense canopies. At Figure 2, one observes that from the RGB image at nadir, it is difficult to differentiate between the canopies even by visual interpretation. However, the altitude information of the canopies suggests the separation from the formation of the saddle point between the top points.

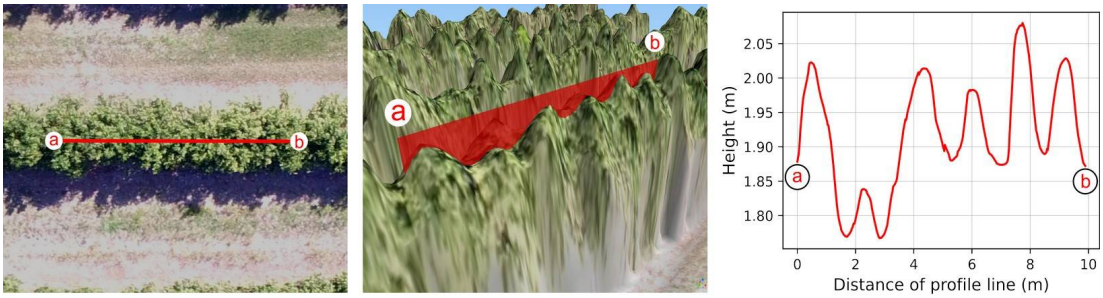

Figure 2. Canopy Height Model (CHM) derived from the photogrammetry process applied to UAV images. (**A**) refers to the true-color composite; (**B**) refers the 3D perpective and, (**C**) a canopy height profile.

2.3. The Mask R-CNN

The detection and delimitation of treetops was performed using the Mask R-CNN [18], one of the main frameworks currently used for instance segmentation in remote sensing images [30]. This mask operates in two modules, one for detecting regions where the presence of the object has been detected and discriminated and the other for segmentation at the pixel level of object boundaries. Its structure (Figure 3) is defined by adding this second module to the structure of the Faster R-CNN [31], a network which can detect objects without defining their borders. The following describes its functioning.

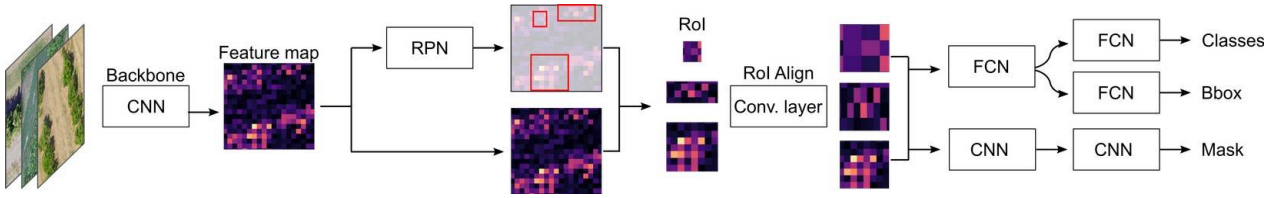

Figure 3. Mask R-CNN Architecture.

The input image is inserted into the backbone structure, which performs a set of convolutional and pooling operations, extracting visual attributes at different scales. In this study, the ResNet101 architecture was used. The attribute maps generated by the backbone are used as input to the Region Proposed Network (RPN), which computes the probable regions containing the desired objects from patterns extracted from the attributes. For each region with a high probability of an object occurrence, multiple anchor boxes (or RoI, Regions of Interest) are computed, depending on the scale and ratio parameters defined for the model. In this study, scales 4, 8, 16, 32, 64 (pixels) and ratios 0.5, 1, 2 were used, resulting in 15 anchor boxes for each region found.

These parameters were defined according to the coverage and proportion of the crowns in the images of different spatial resolutions. Additionally, the same scale interval was maintained despite slightly different values found by [19]. Due to the different sizes and proportions considered for the RoIs, the algorithm performs a homogenization of its sizes, through a convolution layer called RoI-align. RoIs with the same size and scale are used as input in three different processes: (1) the softmax classifier preceded by a Fully Connected Network (FNC) that indicates the class and respective probability of each identified object; (2) the boundary box regressor preceded by an FCN for delimitation of the object's bounding box; and (3) according to the increment of the Faster R-CNN by the

Mask R-CNN, the delimitation of the object's mask, at pixel level, through convolution layers [18].

The CNN R-Mask algorithm was obtained and adapted from the implementation referred by [32]. The adaptation carried out includes the possibility of using images with more than 3 bands and images in TIFF format, preserving the geospatial information, which subsidized the geolocation of the inferred masks. The implementation performed uses COCO-like annotation in JSON format to represent the ground truth required during the training phase.

2.4. Individual Tree Canopy Dataset

To create the ground truth dataset, a methodology was adopted aiming to reduce the time and effort demanded by manual activity for image processing techniques. This methodology provides a possibility of simplifying the process, and it can be used in future works to create a dataset of identification of tree canopies not only with the punctual location or squared borders of the instances but with masks that outline its limits. From mask, the point information and the surrounding rectangle can be derived from the extraction of the centroid and bounding box, respectively.

The proposed methodology for the delineation of the canopies (Figure 4) was performed according to the following sequence:

1. Extraction of the Color Index of Vegetation Extraction (CIVE) [33] whose equation is given by:

$$\text{CIVE} = (0.441\text{“red”}) - (0.811\text{“green”}) + (0.385\text{“blue”}) + 18.78745 \quad (1)$$

   where red, green, and blue are the respective bands of the image;
2. Application of the morphological opening with two different window sizes (3 and 5 pixels) and smoothing in the CIVE image;
3. CIVE image threshold operation from a specific threshold for each image;
4. Conversion of the threshold image from raster to vector (polygon) and selection of the area represented by the set of canopies;
5. Individualization of the polygon representing the set of canopies by manual clipping.

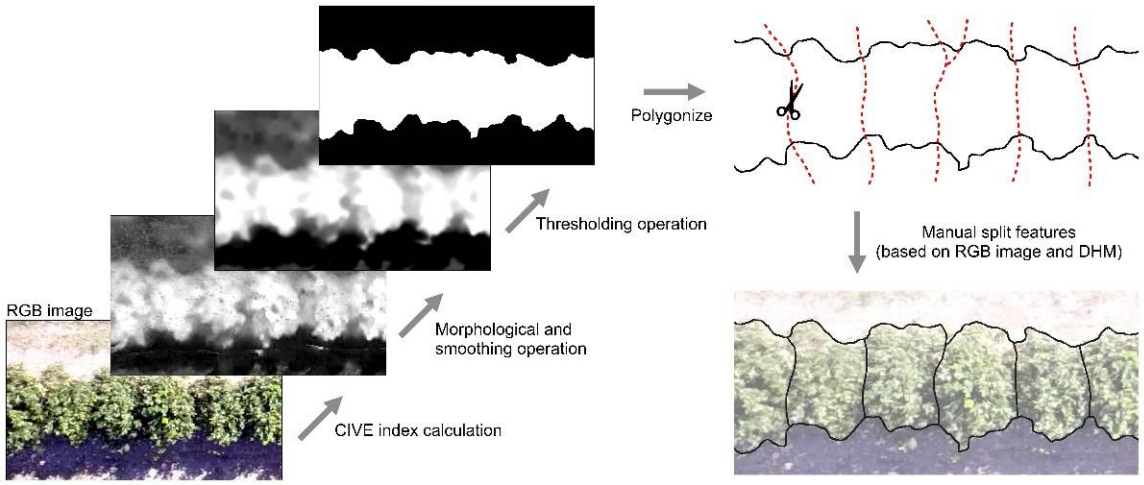

Figure 4. Framework of process adopted for crown delimitation.

As a result of the above sequence, the manual work (Step 5) was only necessary to determine the separation line (or eventually the intersection) between two canopies in regions where there was some level of contact between them, and so manual identification and delimitation became unnecessary.

At the point of contact between two very dense trees, visual identification of the line that separates both can be a difficult and a skewed task if performed only from the RGB image. However, when considering the morphological aspect of the canopies, the occurrence of a saddle point formed by neighboring trees is common, regardless of the proximity between them, which facilitates the individual discrimination. Therefore, for the visual identification of the borders between the canopies, in addition to the RGB image, the CHM was used, which provided information on the relief of the treetops, needed especially in those areas where it is not possible to individualize the trees based on color aspects and texture from the RGB image.

Furthermore, to increase the feasibility of manual canopy delimitation, an algorithm was developed to convert geographic data into a coco-type annotation, the format used in the Mask R-CNN. Thus, all the steps mentioned in this topic were carried out within the GIS environment (QGIS version 3.16) and afterwards, automatically the masks represented by the vector features (polygons) were converted directly into annotations (JSON format) referenced and geo-located in its respective images.

2.5. Training of Models

The pre-processed images were cut into square patches (aspect rate = 1) to be used in the segmentation models. The patches present horizontal and vertical overlap and its size is proportional to the spatial resolution of each original image. In addition to the images, the vector data containing the ground truth with the delimitation of the canopies was also cut following the same cut grid. So, each sample used for training or detection inference corresponds to a patch.

Two segmentation models were defined: the first (RGB model) using RGB images with the three original ortho-mosaic bands (Red, Green and Blue) and the second (RGBC model) using images with 4 bands, namely Red, Green, Blue, and CHM. To homogenize the radiometric resolution between the bands of the RGBC model images, the CHM was converted to 8 bits by a linear transformation. Hence, all the bands of the patches used in each model have a pixel value range between 0 and 255.

The analysis considered the three areas under study indistinctly, since the focus of the work is to evaluate and compare the canopy detection capacity of RGB and RGBC images, considering the spatial complexity and variability of planting characteristics, such as orange variety, age of orchard, presence of vegetation cover, or exposed soil between planting lines and different tree heights and planting densities. Therefore, training, validation of training, and inference were performed considering the three plots uniformly.

In total, 1715 patches were used for each of the two models, with 1031 (60%) used for training, 342 (20%) for training validation, and 342 (20%) for testing. The characteristics adopted in each patch set are shown in Table 2. For all patches in the three fields adopted, the maximum number of instances (canopies or fractions) per patch was 30.

Table 2. Characteristics and number of patches used in both models.

Plot	Dimensions (x per x)		Overlap (px)	GSD (cm)	Area (m^2)	Number of Patches				
	px	m				Train.	Val.	Test	Total	
A	512	11.3	100	2.2	128.7	399	133	133	665	
B	256	18.2	60	7.1	330	196	64	64	324	
C	512	12.8	100	2.5	163.8	436	145	145	726	
						1031	342	342	1715	Total

The two models were similarly trained, with an initial learning rate of 0.001 and a momentum of 0.9. The epoch corresponds to a cycle of use for all training samples, which is formed by iterations (batch size). These in turn may have the same number of samples. Fifty epochs were adopted, each with 1031 iterations and, to increase the learning process,

a 10-fold reduction in the training rate was considered between epochs 15 and 27 (0.0001), 27 and 39 (0.00001), and 39 and 50 (0.000001).

The models were not randomly started during the training process. Both networks were loaded with pre-training parameters of the COCO dataset. However, the parameters loaded correspond to pre-training on RGB images and because the RGBC model is executed with 4-band images, the weights of the first layer for this model were started randomly. Additionally, data augmentation was randomly applied to the training images prior to their entry into the models, and the changes applied to each of the samples correspond to one or two of the following transformations: (1) horizontal mirroring, (2) vertical mirroring, (3) rotation by 90°, 180°, or 270°, and (4) change in the brightness value of pixels in the interval between 50% and 150%.

The evaluation of the Mask R-CNN training and validation is carried out by a set of metrics and each one evaluates a particularity of the detection activity. The metrics used in this study are (1) class loss which indicates how close the model is to the correct class, i.e., canopy or background, (2) bounding box loss—the difference between the bounding box parameters (height and width) of the reference and the inference, (3) mask loss—the pixel-level difference between the reference mask and the inference, and (4) total loss—the sum of all other metrics. After the training phase, the metrics were computed and the learning evolution of the two models is shown in Figure 5.

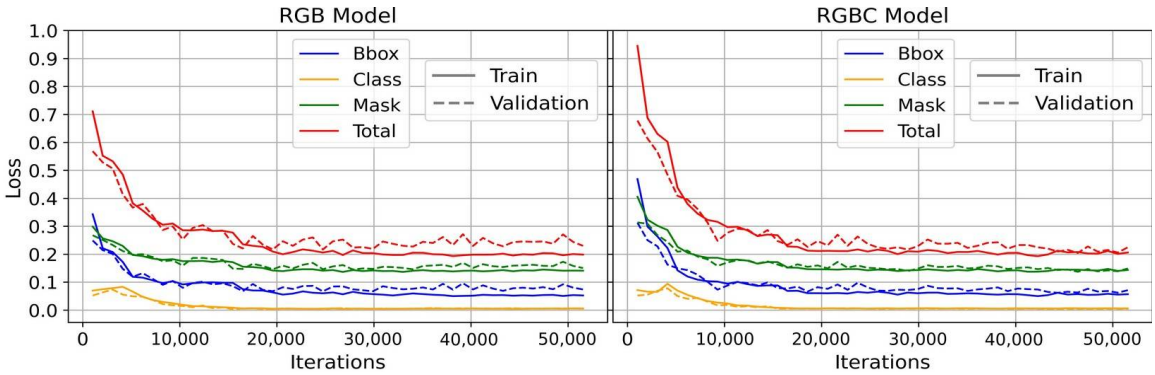

Figure 5. Evolution of validation metrics during Mask R-CNN training.

2.6. Unsupervised Merge Process

Applying the model described for inference, the possible limits of the trees in the original images fractionated by the patches are obtained. However, the overlap between the patches means that some trees may occur redundantly. Redundancies can be a part of a tree located at the end of a patch or an entirely duplicated canopy. Redundancy is identified by the overlapping of two or more masks related to the same tree. In order to prevent it to degrade the detection accuracy, a result aggregation methodology was developed based on the topological relationship between the inferred candidate masks.

The overlap section between patches was defined in such a way that at least one of the generated patches covers fully the treetops in the intersection region. So, for the refinement of the result it is necessary to identify the redundant masks and delete them, and it is not necessary to unite two or more masks to generate a final polygon. Candidate masks which do not overlap with the others are left unchanged.

It is noteworthy that in stands with a high density of plants, the overlap between them can occur naturally and, assuming this possibility, it is not possible to simply exclude the tree segments that overlap. During the analysis, it was observed that the adoption of this procedure implies the exclusion of non-redundant masks and by default the reduction of detection accuracy.

The unsupervised algorithm proposed for this activity consists of the following steps:

1. For each candidate mask, calculate the intersections between it and the others;
2. Evaluate the spatial relationship between intersections and masks and exclude the mask if any of these conditions are met:
   - 2.a Existence of at least one intersection with an area corresponding to minimum 50% from the area of the original mask;
   - 2.b The sum of all intersecting areas of the mask corresponds to at least 50% of its area;
   - 2.c Mask area is less than 1/3 the area of any other mask intersecting with it.

The conditions defined for the refinement were empirically selected from the raw results obtained at the different plots. The natural overlap in the treetops occurs in a subtle way and does not exceed more than 50% of the canopy from each one, either by unilateral (condition 2.1) or bilateral (condition 2.2) overlapping. This overlap occurs between two large trees with similar areas, and it is unlikely, due to the homogeneous spacing adopted in planting. The overlap between a small canopy with a large canopy (condition 2.3) assists in removing false detections, often related to weeds from the plant's surroundings.

The development of this algorithm allowed for the adoption of a unified and automated framework, including all the segmentation procedures of the two previously trained models, which receive as input the ortho-mosaics and the CHMs from pre-processing and the output is the vector data corresponding to the delimitation of the canopies in the entire area defined for inference, without any reference to the division of the patches adopted in the Mask R-CNN (Figure 6). After the inference, the treetops were submitted to the validation of the result from the comparison with the ground truth data extracted manually.

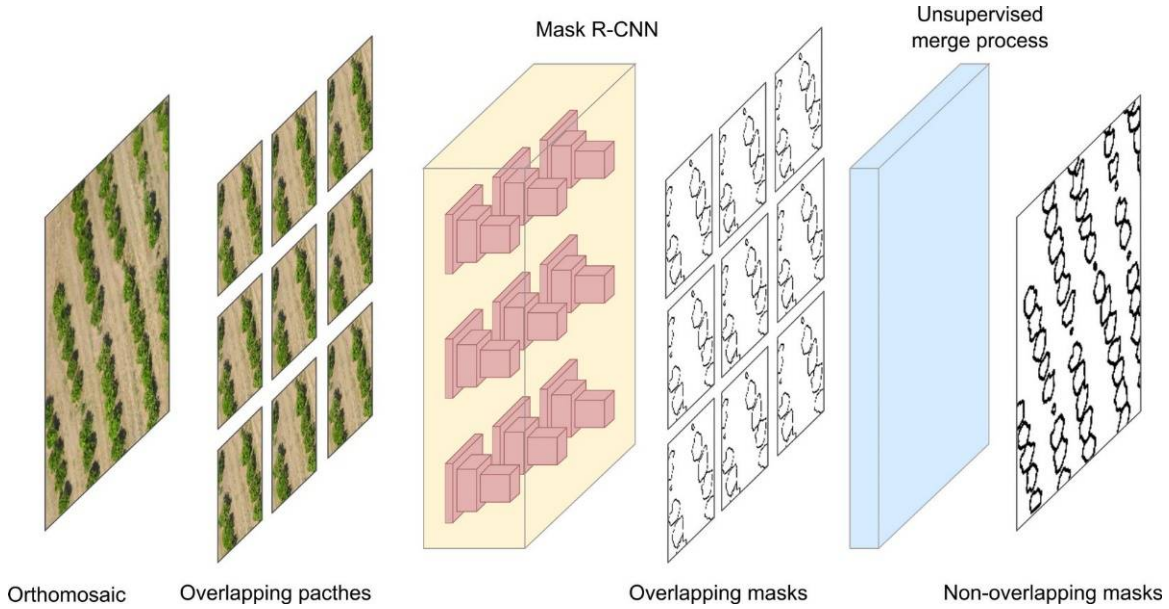

Figure 6. Framework (from left to right) adopted in process of canopy identification and result mosaicking.

2.7. Evaluation Metrics

To validate the proposed method, two evaluation approaches were applied to each of the segmentation models. The first focused on the performance of the individual detection of trees in the test area and the second focused on the precision evaluation of the delineation from each identified canopy. As the metrics of the two approaches have similar names, a suffix was adopted in each to discriminate any case: _det for detection and _del for delineation. In both cases, the validation considered the entire set of inferred and reference canopies.

2.7.1. Number of Detected Trees

All inferred trees were validated through the spatial relationship between their limits and the borders of their respective ground truth by the Intersection of Union (IoU) metric (Figure 7). IoU corresponds to the ratio between the areas of the intersection of the corresponding pair (ground truth and inference) by union. The higher the conformity of the tracings of the two masks, the closer to 1 is the IoU value. A sample was considered true positive (TP) when IoU > 0.5, and to guarantee the integrity of the validation, each reference was used only once when comparing with the inferred canopies.

Hence, three possible results were considered: (1) TP, when the tree was correctly identified; (2) false negatives (FN), when an omission error occurred, i.e., an existing plant was not identified; and (3) false positives (FP), when an inferred mask did not correspond to an existing plant, resulting in a commission error.

From these results, the metrics precision_det (Pdet), recall_det (Rdet), F1-score (Fdet), and overall accuracy (OA) were computed for each of the two models using the following equations:

$$\text{Precision\_det (Pdet)} = \frac{TP}{TP + FP} * 100 \quad (2)$$

$$\text{Recall\_det (Rdet)} = \frac{TP}{TP + FN} * 100 \quad (3)$$

$$\text{F1} - \text{score\_det (Fdet)} = \frac{2 * Pdet * Rdet}{Pdet + Rdet} \quad (4)$$

$$\text{Overall Accuracy (OA)} = \frac{TP}{TP + FN + FP} * 100 \quad (5)$$

2.7.2. Tree Canopy Delineation

The validation for delineation of objects can be made by different methods. In this study more than one approach was adopted for comparison with the existing literature and to suggest future works. The validation included the evaluation of the IoU value itself, the metrics precision_del (Pdel), recall_del (Rdel), F1-score_del (F1del) and the statistical analysis of the canopy area value by linear regression.

The Pdel and Rdel metrics were calculated according to an increment of the approach proposed by [29], who analyzed the relationship between the areas of the bounding boxes of the inferred, the reference segments, and their intersections. The modification proposed in relation to the work mentioned is that the delimiting masks themselves were considered instead of the bounding boxes. Pdel is obtained by the ratio between the intersection area of each corresponding pair and the area of the inferred mask; Rdel is obtained by the ratio of the intersection area over the reference mask; F1-score_del is obtained by equation IV with the metrics Pdel and Rdel, respectively replacing Pdet and Rdet. Figure 7 below displays each metric graphically.

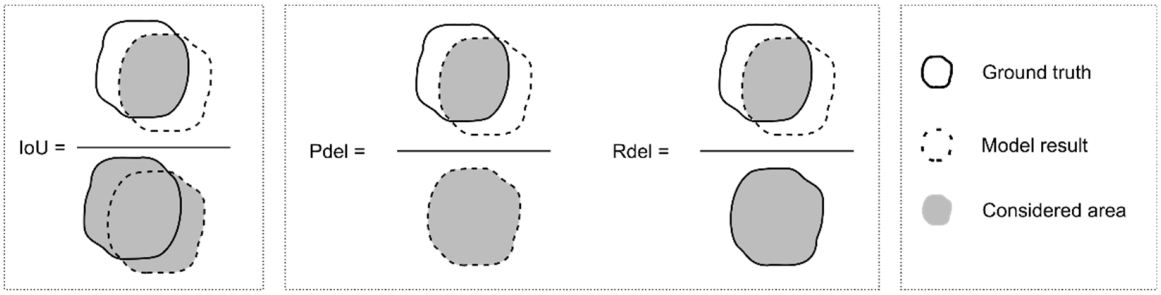

Figure 7. Schematic illustration of delineation metrics.

A regression analysis was performed to evaluate the relation between area and perimeter of the reference canopies and those estimated by the proposed method. For each

analysis, the coefficient of determination (R^2) was computed, which indicates the degree of correlation between the two variables, the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percent Error (MAPE). The higher the R^2, the greater the correlation between the variables and the lower the RMSE, MAE, and MAPE, the closer. The four metrics were used for the analysis.

3. Results

3.1. Number of Detected Trees and Merge Process

The results of the two models regarding the detection of trees are shown in Table 3. The manual identification and delimitation resulted in 2117 canopies in the inference area, used as ground truth. The RGB model correctly identified 2086 canopies, incorrectly 190, and failed to identify 31 trees. The RGBC model had a higher performance in the three situations and correctly identified 2108 crowns, incorrectly 56, and failed to identify 9. All the evaluation metrics for the identification of the RGBC model were superior to those of the RGB model and, disregarding overall accuracy, the largest difference (4.98%) between the models was given by the Precision_det—which evaluates the influence of FP—and the lowest (0.4%) by recall_det—which assesses the influence of FN.

Table 3. Results of the crown detection process.

Metrics	RGB Model	RGBC Model
Detections/Ground Truth	2276/2117	2164/2117
TP	2086	2108
FP	190	56
FN	31	9
Overall Accuracy	90.42%	97.01%
Precision_det	91.65%	97.41%
Recall_det	98.54%	99.57%
F1-Score_det	94.97%	98.48%

After validating the identification, it was found that none of the FN occurrences were caused by the application of the result from the aggregation methodology (patch merge). The redundancies caused by the overlapping image patches were correctly identified and all were suppressed, as well as all non-redundant inferences were kept unchanged in the final result. Figure 8 summarizes some examples of the performance from the clustering process.

The unidentified crowns (FN) by the RGB model were the result, in some cases, of the mistaken identification of only one crown when there were two or more dense trees (under-segmentation), while the RGBC model was able to perform, with some exceptions, the correct identification of the different crown units. This indicates that, in general, the RGB model did not correctly identify the separation line between two contiguous trees, which was not observed with the same frequency in the RGBC model.

As for the FP, the biggest difference between the two models is due to the incorrect identification of crowns located between planting lines because of the vegetation cover. The canopy height information entered in the RGBC model provided a greater ability to identify between what is a real canopy and what is grass. Furthermore, cases of incorrect identification (FP) by the RGB model are also associated, at some level, with the incorrect detection of more than one canopy for the same tree when there are shadows that imply an over-segmentation (Figure 9) of the canopy detection process.

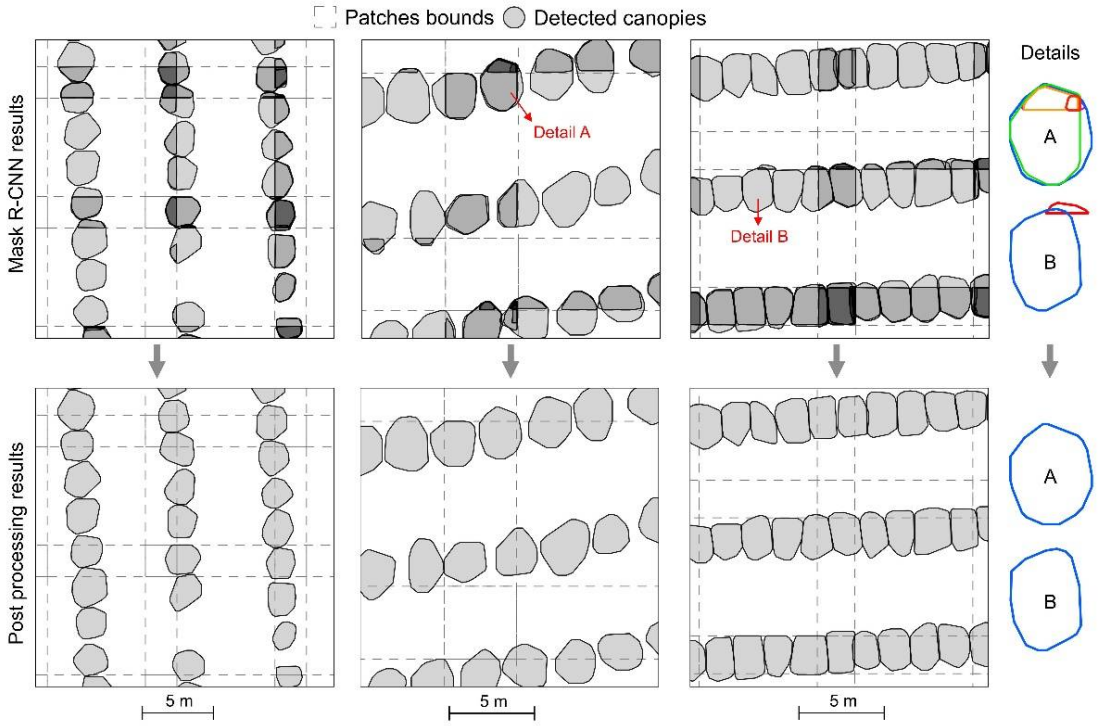

Figure 8. Results of unsupervised merge process.

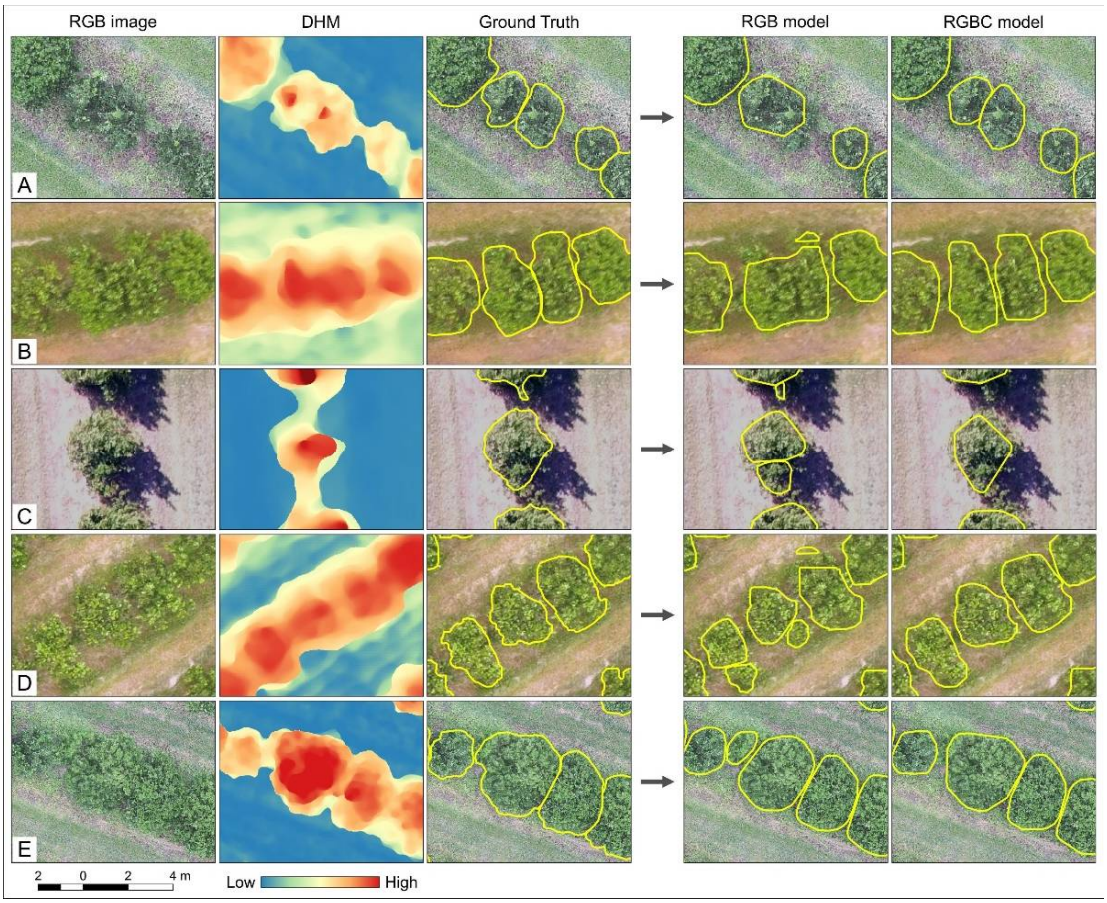

Figure 9. Results of canopy detection and delineation process. (**A**–**E**) indicate different canopy scenarios of orange plantation.

3.2. Tree Canopy Delineation

The evaluation of the design by calculating precision_del, recall_del and f1_del was performed for the crowns correctly identified (IoU > 0.5) in each model and the results are shown in Table 4. These metrics are calculated for each tree and, to evaluate the overall performance, the mean and standard deviation of every one were computed. The RGB model showed an average precision_del of 93.5%, a recall_del of 91.7%, and a F1 of 92.3% among a set of 2086 crowns. The RGBC model showed a positive variation in relation to precision_del, with 96.1%, and a negative variation related to recall_del, of 87.8%, and F1, with 91.6%, in a total of 2108 crowns.

Table 4. Metrics results of delineation process.

Metric	RGB Model	RGBC Model
Precision_del	0.935 ± 0.082	0.961 ± 0.051
Recall_del	0.917 ± 0.058	0.878 ± 0.067
F1-Score_del	0.923 ± 0.060	0.916 ± 0.046

Visually (Figure 10), it is possible to see that a significant difference between the two results is the coverage of the trace from the inferred masks. In the RGB model, the crown masks generally covered the entire region comprised visually by the tree canopy, while in the RGBC model there is a slight retreat from its limits. This is evidenced by the high value of precision_del associated with the low value of recall_del. A possible justification for this result is the difference in the nature of the information obtained from the RGB bands and by the CHM, because while the RGB image with very high spatial resolution is able to represent the extremities of the canopies by the reflectance from even the most extreme leaves, the generation and the accuracy of the CHM is limited by the Structure from Motion process and the filtering performed in the DSM to generate the DTM.

Figure 10. Examples of the difference in canopy delineation between models.

The IoU average was calculated for all the inferred crowns presenting an intersection with some reference and only for those with values above 0.5 (TP), the results are shown in Table 5. In all cases, the RGB model showed higher average values of IoU than the RGBC, with an average of 0.862 considering all crowns and 0.867 only for those with IoU > 0.5. The RGBC model presented 0.847 and 0.848 for all values, as well as for those higher than 0.5.

Table 5. Mean IoU values of the correctly identified crowns.

Model	All	>0.5
RGB	0.862 ± 0.081	0.867 ± 0.061
RGBC	0.847 ± 0.070	0.848 ± 0.066

On the other hand, analyzing the histograms of the two models (Figure 11), one observes that the RGB model presents a maximum IoU value lower than the RGBC one (0.969 and 0.997) and inferred a greater number of crowns that, despite containing an intersection with the respective reference, had a lower IoU than the minimum necessary to be considered a TP (0 < IoU < 0.5). Furthermore, the RGB model resulted in a total of 96.72% (2037) of the crowns with an IoU above 0.7, while the RGBC model reached 97.58% (2060), indicating that, although both models present high fidelity in relation to the ground truth, the RGBC performed slightly better in the delimitation of the traces that agree most with the reference.

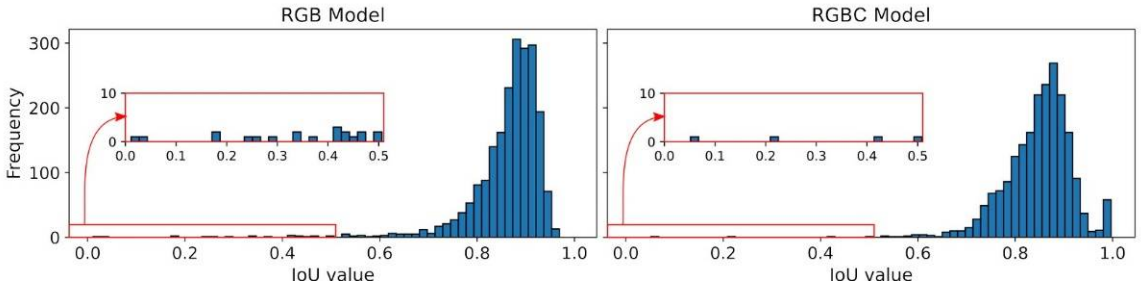

Figure 11. Histogram of IoU values from models.

To determine the relationship of the area and perimeter of each resulting mask with the respective reference, a linear regression analysis was performed and the results are shown in Figures 12 and 13. According to Figure 12, both models achieved results with an expressive correlation between the areas of the corresponding pairs (predicted and reference), and the RGB model has slightly better results than the RGBC. The coefficient of determination (R^2) of the RGB model is 0.973 and 0.967 for the RGBC model. Additionally, the metrics related to the residuals of the RGB model are also slightly better, with RMSE = 0.439 m^2, MAE = 0.320 m^2, and MAPE = 6.90% and RMSE = 0.492 m^2, MAE = 0.357 m^2, and MAPE = 7.98% for the RGBC model.

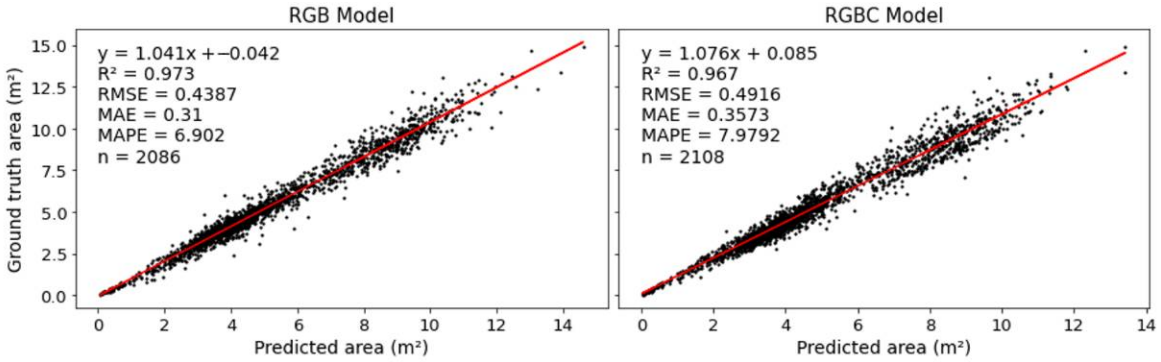

Figure 12. Linear regression results pafor crown area.

Regarding the perimeter (Figure 13), the correlation between the corresponding mask pairs is also high, with R^2 = 0.972 for the RGB model and R^2 = 0.969 for the RGBC model. The small difference between the metrics also occurred in relation to the residuals, with the RGB model results slightly better. The following results were obtained for the RGBC model: RMSE = 0.353 m, MAE = 0.261 m, and MAPE = 3.46% and for the RGB model: RMSE = 0.382 m, MAE = 0.291 m, and MAPE = 3.85%.

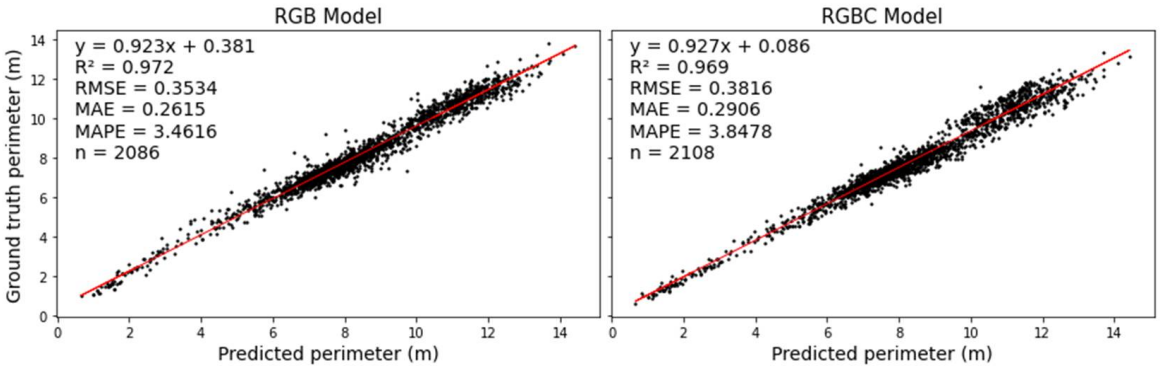

Figure 13. Linear regression results for crown diameter.

4. Discussion

4.1. Tree Canopy Delineation

Recently, the use of Mask R-CNN for the detection and delimitation of targets by remote sensing images has been increasingly used for the study of built [34] and natural environments [14,35] or for agricultural applications [13,19]. Although the agricultural areas present more uniform characteristics than natural environments, the evaluation of this study considered a complex agricultural area, with different planting ages and varieties, spacing and coverage between the planting lines and, above all, with different planting densities, one of the main constraints for the accurate detection of tree canopies [20,25].

In this work, the combined use of RGB images with CHM was made to increase the performance of activity, obtaining results with high accuracy. For the same conditions, the use of CHM significantly increased detection accuracy, if compared to solely using RGB bands (overall accuracy from 90.4% to 97.0% and F1 from 95.0% to 98.5%).

Other studies have also evaluated the use of CHM in target detection with deep learning models. The best combination of bands for detecting Chinese fir canopies using the Mask R-CNN trained with manually delimited ground truth were evaluated by Ref. [13]. The authors evaluated 9 band combinations and concluded that the best performance is obtained with the RGB combination with F1 = 94.7%, however, the RGB + CHM combination was not evaluated, and the CHM was evaluated individually. The use of 150 combinations of attributes from RGB images and DEM to identify treetops in different environments was evaluated by [36]. The best result combination included data obtained from the RGB (grayscale) and DEM (slope, hillshade, CHM) bands, with an F1-score of 92.5%. Comparing the results obtained in this research with previous studies addressing the use of RGB images and DEM-based data, the performance of the proposed RGBC model for detecting tree canopies in agricultural lands with heterogeneous characteristics plots out.

Regarding the detection of canopies using materials and methods slightly different from those adopted in this research, some studies represent the state of the art regarding the detection/counting of plants from different species in agricultural plots. Table 6 summarizes the main findings of these surveys, covering the detection method used, the spectral characteristics demanded (bands), the tree species, the type of output of the result, which are limited to patches or distributed image-wide and presenting the detection/count evaluation metrics.

Among the above listed studies, some of them consider a CNN-based model for the detection of orange tree canopies, obtaining F1-score results above 90%, discussing the influence of planting density on the results obtained. The need for well-defined spacing between neighboring canopies was explained by Ref. [20]. Consequently, the characteristics of the studied plots are different from this study, and they have application limitations. On the other hand, Ref. [21] used images of an orchard with plant density characteristics similar to those of our plot 3 adopted in the present study and, with a detection method by punctual canopy location, reached an F1-score result of 91.1%. These results suggest

that the canopy altitude information, acquired by RGB images and SfM techniques, has the potential to help in the identification of spatially dense canopies.

Table 6. Characteristics and results of research similar to this study.

Research	Method	Bands Used	Specie of Trees	Output Type	Output Bounds	Precision	Recall	F1
[13]	Marker-Controlled Watershed	CHM	Chinese fir	Mask	Image	0.770	0.972	0.859
[21]	CNN-based	RGB	Corn	Point	Patch-based	0.856	0.905	0.876
[13]	Local Maxima	CHM	Chinese fir	Point	Image	0.793	0.985	0.879
[21]	CNN-based	RGB	Orange	Point	Patch-based	0.922	0.905	0.911
[13]	Mask R-CNN	RGB	Chinese fir	Mask	Patch-based	0.957	0.937	0.947
This study (Existing approach)	Mask R-CNN	RGB	Orange	Mask	Image	0.917	0.985	0.950
[25]	CNN-based	Multispectral	Orange	Point	Patch-based	0.950	0.960	0.950
[20]	CNN-based (before post refinement)	Multispectral	Orange	Bbox	Image	0.987	0.981	0.984
This study (Proposed approach)	Mask R-CNN	RGB + CHM	Orange	Mask	Image	0.974	0.996	0.985
[15]	Local Maxima + Marker-Controlled Watershed	RGB + CHM	Apple Pear	Mask Mask	Image Image	0.997 0.995	0.983 0.990	0.990 0.993

In our study, four characteristics stand out from previous works with similar results, namely: (1) the proposed technique is based on low-cost UAV remote sensing systems, since it uses only RGB and CHM images derived from less costly sensors. When compared to multispectral sensors, (2) the plots studied have a significant variability in spatial and spectral characteristics, allowing the identification of trees with a higher density, due to the altitude and canopy structure, delivered by the CHM; (3) our result include the entire area of the image analyzed, without fractional analysis or manual post-processing; and (4) the identification result does not occur only with the punctual location or with the surrounding rectangle, but with the delimiting mask of the crown limits, subsidizing analyses with other approaches besides counting plants.

The results presented by Ref. [15] are slightly superior to those obtained in our study. However, the methodology applied by these authors demands a selection of unique detection thresholds for each analysis. Furthermore, adopting in this work plots with distinct spatial and spectral characteristics, it is necessary to have the analyst's prior knowledge or a trial-and-error approach to define the parameters of each. This condition reduces the feasibility of the analysis and causes difficulty in the development of automated detection frameworks. Contrarily, the use of the previously trained Mask R-CNN shows a potential for detection in heterogeneous plots with very similar results.

4.2. Tree Canopy Delineation

The tree canopy detection proposed in this study has some degree of dependence on the correct individual detection of them, since over-segmentation and under-segmentation is already a factor of inconsistency for the individual design. The RGBC model showed a better performance when it correctly identified a larger number of individual canopies. However, the results of the design metrics, the area correlation, and the perimeter between the inferred and reference canopies shows that, in general, the delimitation performed by the RGBC model is lower than the RGB model. The difference between the tracings

is almost unanimously due to under-segmentation and some points can be mentioned to justify this result.

Initially, during the generation of the point cloud which originates the DEM through the SfM-MVS technique, the complexity of the vegetation structure is naturally a challenge [37], since the irregular layout of the crown edge causes a smoothing in the CHM, reducing its coverage in relation to the flat area occupied by the plant [38]. Secondly, it is noteworthy that all images are acquired without control points, which influences directly the geometric quality of the UAV images to generate the DEM and consequently the CHM and the ortho-mosaic. Thirdly, the images were acquired from a flight plan that contemplated only the nadir sight and the longitudinal and lateral overlap of the images. It is possible that the CHM reconstitution process is more accurate if a flight plan with different views and crossed image lines is adopted, increasing the overlap between the scenes.

Although it is possible to improve the CHM generation, the RGBC model shows promising results in crown detection and it is comparable with the results of previous studies. As per Ref. [15], using Local Maxima and Marker-Controlled Watershed techniques used to segment tree canopies in agricultural regions, the best results for apple plants were obtained: R^2 = 0.87, RMSE = 0.72 and MAE = 0.57 for the correlation between the estimated and reference crown area and R^2 = 0.81, RMSE = 0.48, MAE = 0.39 for the perimeter. For the detection of dense canopies in a forest environment using the Mask R-CNN, R^2 = 0.93 was obtained in the correlation analysis of the inferred canopy areas with those in the reference as per Ref. [29]. A CNN-based model to delimit individual trees in apple plantations was used by Ref. [39]. The results of the correlation coefficient between the area and perimeter of the results obtained with the reference were R^2 = 0.80 and R^2 = 0.79 for area and perimeter, respectively. Regarding these studies, the results with the proposed method are extremely viable to increase the viability and precision of tree canopy delimitation in the agricultural environment, which facilitates those studies involving the structural characteristics of each plant in the plot.

4.3. Merging of Patches

For the detection and delimitation of treetops, a Mask R-CNN was used, whose input are patches of images with significantly smaller areas (limit of 256 × 256 or 512 × 512 pixels) than the complete field. This approach is common in studies involving object detection or instance segmentation in remote sensing images [13,15,19,20,25], as they are representations with variable dimensions and are generally superior to conventional images. However, unlike the studies mentioned, in our study an algorithm for automated mosaicking of the results is proposed, which presents a significant performance and eliminates the need for manual post-processing of the results obtained in each image fraction.

There are few studies that address some similar processing for applications in instance segmentation. Ref. [29] proposed a similar approach for detecting crowns in dense forests, joining any two trees at the edges of patches with polygonal intersections among them. Ref. [28] proposed a moving window mosaicking method applied in an analogous region as the overlapping of the patches in this study. From a modification of the non-maximal suppression algorithm, these authors perform the deletion of redundant objects with a methodology similar to condition 2.3 of the proposed algorithm but considering only the bounding box of the masks. Nevertheless, these techniques are not robust enough to contemplate the possibility of a real intersection in two neighboring crowns and, above all, the filtering of large false positives that overlap more than one real crown (filtered by condition 2.2 of the proposed approach), would increase errors by omission and commission, respectively, if they were applied for the same purpose of this study. Therefore, the proposed method has a strong potential to allow the analysis focused on detection and delimitation of canopies in plots with high planting density—where the overlap between canopies is natural from images of large areas applied to the Mask R-CNN architecture.

5. Conclusions

In this study, we evaluated the use of images captured by UAVs and photogrammetry techniques using the SfM algorithm for the identification and delineation of treetops located in different spatial densities using deep learning. The architecture of the model adopted, the Mask R-CNN, corresponds to the state-of-the-art in the instance segmentation process and the images used considered exclusively the RBG sensors of digital cameras. Additionally, a methodology for the elaboration of the ground truth was adopted, which increased the viability of the analysis performed. An unsupervised algorithm was developed that contributes to the automation of instance segmentation in remote sensing images.

Therefore, the main contributions of the proposed approach in this study are: (1) the effort reduction for the elaboration of the ground truth and of the training samples; (2) proposal of a model to identify and delimit dense tree canopies in images of different characteristics with high accurate results (Detection: Overall accuracy and F1-score 97.01% and 98.48%, respectively, and design: IoU > 0.5 average of 0.848 and average F1-score of 91.6%); (3) development of a methodology to encapsulate all procedures adopted, reducing the need for manual operation.

The analysis performed did not cover all the possibilities involving the improvement of the instance segmentation process using SfM-based data. Future works will be carried out to explore other agriculture and forestry crops, where the crowns also present high spatial density, using data with different imaging characteristics.

Author Contributions: Conceptualization, F.L., F.M.B. and H.K.; methodology, F.L.; software, F.L.; validation, F.L.; formal analysis, F.L.; investigation, F.L.; resources, F.M.B. and H.K.; data curation, F.L. and F.M.B.; writing—original draft preparation, F.L.; writing—review and editing, F.M.B. and H.K.; visualization, F.L.; supervision, F.M.B. and H.K.; project administration, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)—Finance Code 001.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, *327*, 828–831. [CrossRef] [PubMed]
2. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
3. Leiva, J.N.; Robbins, J.; Saraswat, D.; She, Y.; Ehsani, R. Evaluating remotely sensed plant count accuracy with differing unmanned aircraft system altitudes, physical canopy separations, and ground covers. *J. Appl. Remote Sens.* **2017**, *11*, 36003. [CrossRef]
4. Mohan, M.; Silva, C.A.; Klauberg, C.; Jat, P.; Catts, G.; Cardil, A.; Hudak, A.T.; Dia, M. Individual tree detection from unmanned aerial vehicle (UAV) derived canopy height model in an open canopy mixed conifer forest. *Forests* **2017**, *8*, 340. [CrossRef]
5. Hunt, E.R.; Daughtry, C.S.T. What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *Int. J. Remote Sens.* **2018**, *39*, 5345–5376. [CrossRef]
6. Fan, Z.; Lu, J.; Gong, M.; Xie, H.; Goodman, E.D. Automatic Tobacco Plant Detection in UAV Images via Deep Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 876–887. [CrossRef]
7. Wu, J.; Yang, G.; Yang, X.; Xu, B.; Han, L.; Zhu, Y. Automatic counting of in situ rice seedlings from UAV images based on a deep fully convolutional neural network. *Remote Sens.* **2019**, *11*, 691. [CrossRef]
8. Dadras Javan, F.; Samadzadegan, F.; Seyed Pourazar, S.H.; Fazeli, H. UAV-based multispectral imagery for fast Citrus Greening detection. *J. Plant Dis. Prot.* **2019**, *126*, 307–318. [CrossRef]
9. Pádua, L.; Vanko, J.; Hruška, J.; Adão, T.; Sousa, J.J.; Peres, E.; Morais, R. UAS, sensors, and data processing in agroforestry: A review towards practical applications. *Int. J. Remote Sens.* **2017**, *38*, 2349–2391. [CrossRef]
10. Marques, P.; Pádua, L.; Adão, T.; Hruška, J.; Peres, E.; Sousa, A.; Sousa, J.J. UAV-based automatic detection and monitoring of chestnut trees. *Remote Sens.* **2019**, *11*, 855. [CrossRef]
11. Di Gennaro, S.F.; Matese, A. Evaluation of novel precision viticulture tool for canopy biomass estimation and missing plant detection based on 2.5D and 3D approaches using RGB images acquired by UAV platform. *Plant Methods* **2020**, *16*, 91. [CrossRef] [PubMed]

12. Hao, Z.; Lin, L.; Post, C.J.; Mikhailova, E.A.; Li, M.; Chen, Y.; Yu, K.; Liu, J. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 112–123. [CrossRef]
13. Yu, K.; Hao, Z.; Post, C.J.; Mikhailova, E.A.; Lin, L.; Zhao, G.; Tian, S.; Liu, J. Comparison of Classical Methods and Mask R-CNN for Automatic Tree Detection and Mapping Using UAV Imagery. *Remote Sens.* **2022**, 14. [CrossRef]
14. Chadwick, A.; Goodbody, T.; Coops, N.; Hervieux, A.; Bater, C.; Martens, L.; White, B.; Röeser, D. Automatic delineation and height measurement of regenerating conifer crowns under leaf-off conditions using uav imagery. *Remote Sens.* **2020**, *12*, 4104. [CrossRef]
15. Dong, X.; Zhang, Z.; Yu, R.; Tian, Q.; Zhu, X. Extraction of information about individual trees from high-spatial-resolution uav-acquired images of an orchard. *Remote Sens.* **2020**, *12*, 133. [CrossRef]
16. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
17. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
18. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
19. Machefer, M.; Lemarchand, F.; Bonnefond, V.; Hitchins, A.; Sidiropoulos, P. Mask R-CNN refitting strategy for plant counting and sizing in uav imagery. *Remote Sens.* **2020**, *12*, 3015. [CrossRef]
20. Ampatzidis, Y.; Partel, V. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sens.* **2019**, *11*, 410. [CrossRef]
21. Osco, L.P.; Arruda, M.D.S.D.; Gonçalves, D.N.; Dias, A.; Batistoti, J.; de Souza, M.; Gomes, F.D.G.; Ramos, A.P.M.; Jorge, L.A.D.C.; Liesenberg, V.; et al. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 1–17. [CrossRef]
22. Shao, L.; Tian, Y.; Bohg, J. ClusterNet: 3D Instance Segmentation in RGB-D Images. *arXiv* **2018**, arXiv:1807.08894.
23. Aitelkadi, K.; Outmghoust, H.; Laarab, S.; Moumayiz, K.; Sebari, I. Detection and Counting of Fruit Trees from RGB UAV Images by Convolutional Neural Networks Approach. *Adv. Sci. Technol. Eng. Syst. J.* **2021**, *6*, 887–893. [CrossRef]
24. Mattos, A.B.; Zortea, M.; Macedo, M.M.G.; Ruga, B.C.; Gemignani, B.H. Automatic Citrus Tree Detection from UAV Images based on Convolutional Neural Networks Intravascular Optical Coherence Tomography image analysis View project Automatic Citrus Tree Detection from UAV Images based on Convolutional Neural Networks. In Proceedings of the 31st Conference on Graphics, Patterns and Images: SIBGRAPI 2018, Foz do Iguaçu, Brazil, 29 October–1 November 2018. Available online: https://www.researchgate.net/publication/329240331 (accessed on 20 January 2021).
25. Osco, L.P.; Arruda, M.D.S.D.; Marcato Junior, J.; da Silva, N.B.; Ramos, A.P.M.; Moryia, A.S.; Imai, N.N.; Pereira, D.R.; Creste, J.E.; Matsubara, E.; et al. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 97–106. [CrossRef]
26. Rodrigues, R. *Detection of Sugarcane Crop Rows from UAV Images Using Semantic Segmentation and Radon Transform*; Universidade Federal de Uberlândia: Uberlândia, Brazil, 2020.
27. Garza, B.N.; Ancona, V.; Enciso, J.; Perotto-Baldivieso, H.L.; Kunta, M.; Simpson, C. Quantifying citrus tree health using true color UAV images. *Remote Sens.* **2020**, *12*, 170. [CrossRef]
28. Carvalho, O.; Júnior, O.D.C.; Albuquerque, A.; Bem, P.; Silva, C.; Ferreira, P.; Moura, R.; Gomes, R.; Guimarães, R.; Borges, D. Instance segmentation for large, multi-channel remote sensing imagery using mask-RCNN and a mosaicking approach. *Remote Sens.* **2021**, *13*, 39. [CrossRef]
29. Braga, J.R.G.; Peripato, V.; Dalagnol, R.; Ferreira, M.P.; Tarabalka, Y.; Aragão, L.E.O.C.; Velho, H.F.D.C.; Shiguemori, E.H.; Wagner, F.H. Tree crown delineation algorithm based on a convolutional neural network. *Remote Sens.* **2020**, *12*, 1288. [CrossRef]
30. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
32. Waleed, A. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. GitHub Repository. 2017. Available online: https://github.com/matterport/Mask_RCNN (accessed on 19 June 2020).
33. Kataoka, T.; Kaneko, T.; Okamoto, H.; Hata, S. Crop growth estimation system using machine vision. *IEEE/ASME Int. Conf. Adv. Intell. Mechatron.* **2003**, *2*, 1079–1083. [CrossRef]
34. Wu, Q.; Feng, D.; Cao, C.; Zeng, X.; Feng, Z.; Wu, J.; Huang, Z. Improved mask r-cnn for aircraft detection in remote sensing images. *Sensors* **2021**, *21*, 2618. [CrossRef] [PubMed]
35. Ullo, S.L.; Mohan, A.; Sebastianelli, A.; Ahamed, S.E.; Kumar, B.; Dwivedi, R.; Sinha, G.R. A New Mask R-CNN-Based Method for Improved Landslide Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3799–3810. [CrossRef]
36. Pleşoianu, A.I.; Stupariu, M.S.; Şandric, I.; Pătru-Stupariu, I.; Drăguţ, L. Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model. *Remote Sens.* **2020**, *12*, 2426. [CrossRef]

37. Carrivick, J.L.; Smith, A.M.W.; Quincey, D.J. *Structure from Motion in the Geosciences*; John Wiley & Sons: Hoboken, NJ, USA, 2016; Volume 84.
38. Jayathunga, S.; Owari, T.; Tsuyuki, S. Evaluating the performance of photogrammetric products using fixed-wing UAV imagery over a mixed conifer-broadleaf forest: Comparison with airborne laser scanning. *Remote Sens.* **2018**, *10*, 187. [CrossRef]
39. Wu, J.; Yang, G.; Yang, H.; Zhu, Y.; Li, Z.; Lei, L.; Zhao, C. Extracting apple tree crown information from remote imagery using deep learning. *Comput. Electron. Agric.* **2020**, *174*, 105504. [CrossRef]