

Article

Extraction of Olive Crown Based on UAV Visible Images and the U²-Net Deep Learning Model

Zhangxi Ye ¹, Jiahao Wei ¹, Yuwei Lin ¹, Qian Guo ¹, Jian Zhang ¹, Houxi Zhang ^{1,2,3,*}, Hui Deng ⁴ and Kaijie Yang ¹

¹ Forestry College, Fujian Agriculture and Forestry University, Fuzhou 350002, China; yejance@fafu.edu.cn (Z.Y.); 3205320193@stu.fafu.edu.cn (J.W.); linyuwei@fafu.edu.cn (Y.L.); guoqian@fafu.edu.cn (Q.G.); zhangjian2020@fafu.edu.cn (J.Z.); 000q820060@fafu.edu.cn (K.Y.)

² Key Laboratory of State Forestry and Grassland Administration for Soil and Water Conservation in Red Soil Region of South China, Fuzhou 350002, China

³ Cross-Strait Collaborative Innovation Center of Soil and Water Conservation, Fuzhou 350002, China

⁴ College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China; dengh@cdut.edu.cn

* Correspondence: zhanghouxi@fafu.edu.cn

Abstract: Olive trees, which are planted widely in China, are economically significant. Timely and accurate acquisition of olive tree crown information is vital in monitoring olive tree growth and accurately predicting its fruit yield. The advent of unmanned aerial vehicles (UAVs) and deep learning (DL) provides an opportunity for rapid monitoring parameters of the olive tree crown. In this study, we propose a method of automatically extracting olive crown information (crown number and area of olive tree), combining visible-light images captured by consumer UAV and a new deep learning model, U²-Net, with a deeply nested structure. Firstly, a data set of an olive tree crown (OTC) images was constructed, which was further processed by the ESRGAN model to enhance the image resolution and was augmented (geometric transformation and spectral transformation) to enlarge the data set to increase the generalization ability of the model. Secondly, four typical subareas (A–D) in the study area were selected to evaluate the performance of the U²-Net model in olive crown extraction in different scenarios, and the U²-Net model was compared with three current mainstream deep learning models (i.e., HRNet, U-Net, and DeepLabv3+) in remote sensing image segmentation effect. The results showed that the U²-Net model achieved high accuracy in the extraction of tree crown numbers in the four subareas with a mean of intersection over union (IoU), overall accuracy (OA), and F1-Score of 92.27%, 95.19%, and 95.95%, respectively. Compared with the other three models, the IoU, OA, and F1-Score of the U²-Net model increased by 14.03–23.97 percentage points, 7.57–12.85 percentage points, and 8.15–14.78 percentage points, respectively. In addition, the U²-Net model had a high consistency between the predicted and measured area of the olive crown, and compared with the other three deep learning models, it had a lower error rate with a root mean squared error (RMSE) of 4.78, magnitude of relative error (MRE) of 14.27%, and a coefficient of determination (R^2) higher than 0.93 in all four subareas, suggesting that the U²-Net model extracted the best crown profile integrity and was most consistent with the actual situation. This study indicates that the method combining UVA RGB images with the U²-Net model can provide a highly accurate and robust extraction result for olive tree crowns and is helpful in the dynamic monitoring and management of orchard trees.

Keywords: UAV-based remote sensing; visible-light image; individual tree segmentation; deep learning; tree crown extraction

Citation: Ye, Z.; Wei, J.; Lin, Y.; Guo, Q.; Zhang, J.; Zhang, H.; Deng, H.; Yang, K. Extraction of Olive Crown Based on UAV Visible Images and the U²-Net Deep Learning Model. *Remote Sens.* **2022**, *14*, 1523. <https://doi.org/10.3390/rs14061523>

Academic Editors: Mauro Maesano and Federico Valerio Moresi

Received: 5 February 2022

Accepted: 15 March 2022

Published: 21 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The olive tree is an essential asset with considerable economic, ecological, and cultural value [1]. Its fruit contains more than 17 different amino acids required by the human body [2], and it is widely employed in daily use, medicine, the chemical industry, and

other industries. Automatic monitoring of olive tree growth characteristics and health status is vital for water and fertilizer management, pest control, and fruit yield prediction [3,4]. With the ongoing development of planting scale in recent years, accurate, and effective olive tree monitoring has been necessary. As an essential part of trees, the crown is a crucial indicator of tree growth and health [5] and affects photosynthesis, transpiration, nutrient absorption, and other processes that regulate tree growth and development [6]. Through dynamic monitoring of olive tree crown, information, such as crop growth, biotic stress, and abiotic stress, can be obtained thereby providing important support for agricultural tasks such as irrigation and soil fertilization, branch pruning, pest control, and yield prediction; it can also help fruit farmers carry out scientific planning and decision-making to achieve precise management of orchards. In addition, the single tree information of olive is an essential basis for analyzing fruit tree planting density, spatial variation of individual productivity, and its growth environment. It is also an important indicator for farmers' planting subsidy applications, orchard asset evaluations, and loss assessments due to the fact of artificial or natural special events (such as fires, insect pests, or other natural disasters). Therefore, accurate and timely acquisition of olive tree crown information is of great significance for achieving orchard automation, informatization, and precise management [7].

The traditional way to acquire forest crown information is manual measurement in the field. Due to the measurement conditions and cost, measuring the forest crown in the whole region is challenging; thus, some of them need to be picked as representatives for estimation. However, usually, only easily accessible and limited samples are collected [8], which may lead to a significant bias and certain inaccuracies [9]. Moreover, its application in large-scale measurement is severely limited because of its low efficiency. It is critical to explore a low-cost and efficient method for obtaining accurate forest crown information in this context. In recent years, remote sensing technology has made significant advances and has been proved to be an effective means for resource monitoring [10–15] such as water [16], crop [17], and forest [18]. Although traditional remote sensing (e.g., satellite and aerospace) is effective in monitoring resources on a large scale, as it is prone to clouds. Therefore, it is usually difficult to obtain qualified images on cloudy days. Furthermore, its low spatial and temporal resolution is usually significantly limited in accurately identifying small ground objects, especially those at a single-wood level. Instead, as a new low-altitude remote sensing platform, unmanned aerial vehicles (UAVs) with substantial flexibility can reach a forest that is difficult to enter manually, and it can obtain ultra-high-resolution digital orthophoto maps (DOMs) and digital elevation models (DEMs) with a concise return cycle at the single-tree level that are superior to traditional remote sensing platforms at a small spatial scale [19–22]. Generally, consumer UAVs equipped with a visible RGB sensor have become critical information acquisition tools in the forestry field because of their low cost and good portability [23–25].

Over the past few years, deep learning (DL) techniques have grown in popularity for object detection or segmentation in any imagery (e.g., bio-medical or remote sensing images) [26]. DL approaches are also widely used in remote sensing for typical image segmentation and classification [27,28]. Furthermore, DL-based image segmentation has been particularly applied to UAV data because of its good ability to extract deeper features from high-resolution images. The combination of UAV and DL has realized the rapid acquisition and analysis of forestry information and significantly promoted the intelligent monitoring and management of forestry [6]. Numerous studies about forestry resource investigation and biomass estimation based on DL and UAV techniques have been published. Oscoetal et al. [29] proposed a new convolutional neural network to estimate the number of citrus trees using UAV multispectral images. Ferreira et al. [30] used a deep learning model, DeepLabv3+, to classify palm trees and extract tree crowns in the Amazon basin, which performed well. Lou et al. [31] used UAV and deep learning to identify the crown and width of two pine plantations in eastern Texas and achieved an R^2 of up to 0.94.

However, the mainstream deep learning models still have some shortcomings in the accurate extraction of tree crown information, because the backbone used by these models to extract the global semantic details of the original image are mainly structures such as VGG [32], ResNet [33], and DensNet [34]. Nonetheless, these network structures are mainly designed for image classification tasks. Thus, they easily ignore rich low-level and intermediate semantic features, resulting in reduced ground object extraction completeness. Moreover, these deep learning models, applying to ground object extraction in remote sensing, often need to be pretrained using sizeable remote sensing data sets, such as WHU-RS19 [35], AID [36], and NWPU VHR-10 [37], and then can be applied to small-scale ground object extraction using transfer learning. The overall extraction effect is relatively poor if the features of the target object are vastly different from these data sets. Therefore, it is necessary to design a new deep learning model that does not depend on pretraining weights and has better extraction results than existing models. Qin et al. [38] proposed a new model, U²-Net, a U-shaped deep network model with two nested layers, to address the issues of the traditional models mentioned above. The new model can obtain high accuracy without pretraining a large amount of data. It has a simple structure and good performance in terms of operational efficiency and resource consumption, which provides an excellent opportunity to use UAV images to extract tree crown information accurately. At present, the U²-Net model is mainly used in people's image processing and bio-medical image segmentation. However, few studies combine UAV visible images with U²-Net for tree crown extraction, so it is critical to explore the feasibility and effectiveness of this model for olive tree crown detection.

The objective of this study was to investigate the potential of low-cost UAV in olive tree crown extraction combined with the semantic segmentation model U²-Net. For this purpose, we first constructed an olive tree crown detection and segmentation (OTC) data set from UAV images. Based on this, we undertook the following work: (1) evaluated the reliability of the U²-Net model in the crown information extraction; (2) compared the U²-Net model with other deep learning models to validate its performance. The U²-Net model is expected to provide a cheap and efficient method for crown information extraction to guide economic forest management.

2. Methods

2.1. Study Area

An olive grove (118°57' E, 26°12' N) in Minhou County, Fujian Province, China, was selected as the study area (Figure 1), and there were 751 olive trees. Minhou County, which has an altitude varying from approximately 87 m to 135 m above sea level and has a typical subtropical monsoon climate with sufficient sunshine and abundant rainfall, is located in the lower Minjiang River. The average annual sunshine duration is 1700–1980 h, the average annual precipitation is 900–2100 mm, the average annual temperature is 20–25 °C, the average annual relative humidity is 75–80%, and the frost-free period is 240–320 days, forming a unique and friendly natural environment for olive growth. The total planting area in this region is more than 40 km², and the output is about 50,000 tons/year. This olive planting area and production rank first in Fujian Province, and is known as the “hometown of the olive in China”.

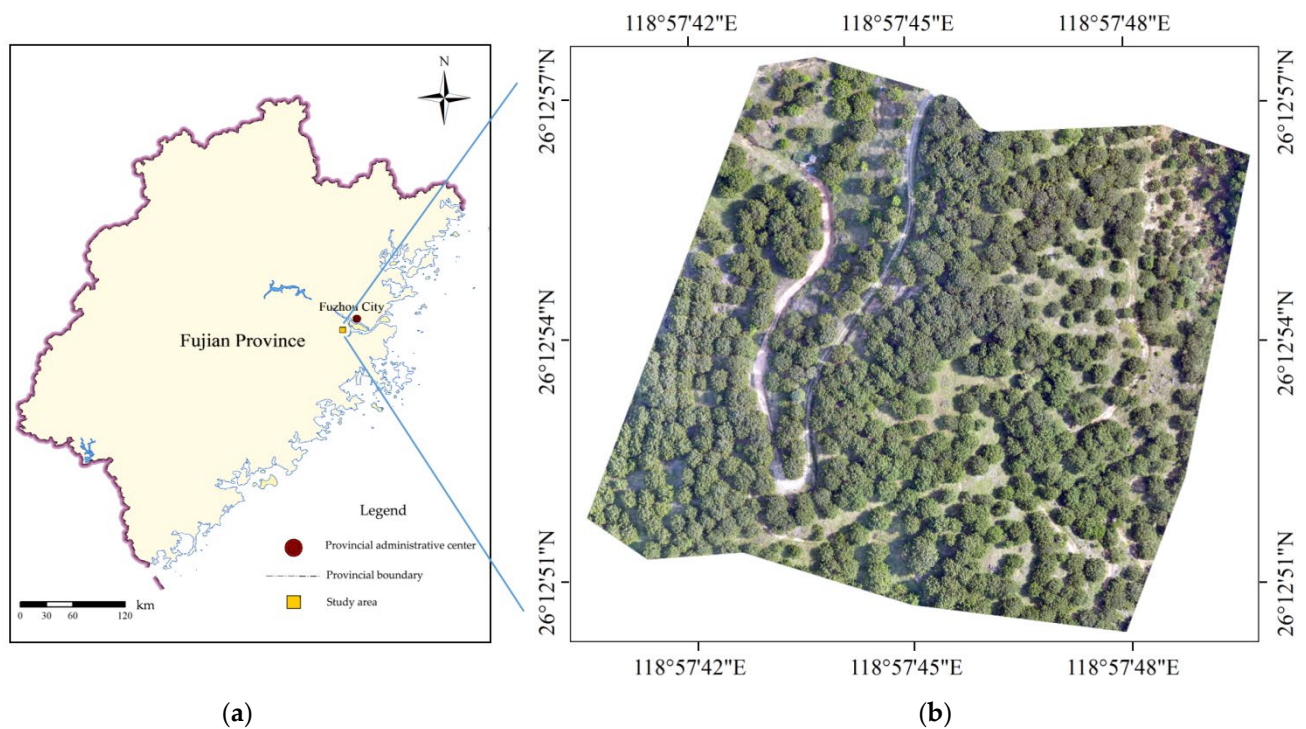


Figure 1. Geographic location of the study area: (a) administrative map of Fujian Province, China; (b) orthophoto of the study area taken by UAV.

2.2. UAV Image Acquisition and OTC Data Set Construction

The UAV images were collected in July 2021 using a DJI Phantom 4 Multispectral (DJI Technology Co., Ltd., Shenzhen, China) equipped with six CMOS (1/2.9-inch), including one RGB sensor for visible-light imaging and five monochrome sensors for multispectral imaging. In order to explore the application effect of a consumer UAV equipped with an RGB sensor on olive crown extraction, we only used an RGB sensor in this study. The sensor had an equivalent focal length of 40 mm and an aperture of $f/2.2$ that could generate an image with a resolution of 1600×1300 pixels. We chose a day with sunny and breezeless weather conditions to perform UAV flying to obtain clear images. We used the remote control automatic flight method for drone aerial photography, and the flight mission planning was set with the DJI GS PRO (DJI Technology Co., Ltd., Shenzhen, China). The flight altitude was set to 30 m, the photos we took had 75% overlap in both heading and sideways, and a total of 1011 photos were collected in the study area. The acquired images were mosaicked using DJI Terra software (DJI Technology Co., Ltd., Shenzhen, China) to generate a visible digital orthophoto map (DOM) with a spatial resolution of 2 cm. The process of constructing the data set is shown in Figure 2. Firstly, the orthophoto images of the whole study area were divided into 110 subimages (each with $1200 \text{ pixels} \times 1200 \text{ pixels}$), and from these, 87 subimages containing olive trees were selected. To make the U²-Net deep learning model learn the spectral features of the olive crown in the visible images more effectively, we used the LabelMe 3.3.6 image annotation software to label the olive crown in each subimage manually.

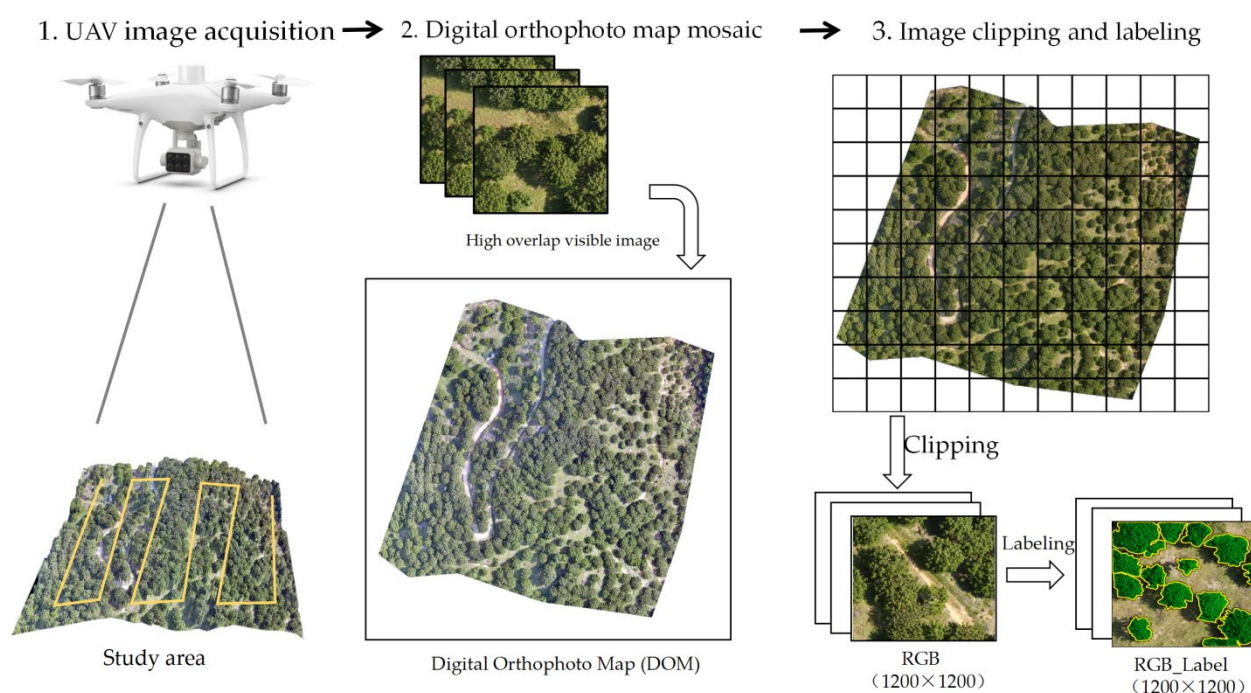


Figure 2. Flow chart of data set's construction.

2.3. Super-Resolution Reconstruction

UAV images with low spatial resolution often have the characteristics of weak contrast, weak texture, and insignificant contour information, and many essential features are often lost after a series of operations of convolution and pooling of deep learning networks. To solve such problems, Wang et al. [39] proposed an adversarial network-based image super-resolution reconstruction method called Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). It has excellent image enhancement performance and can effectively improve the spatial resolution of remote sensing images [40]. Therefore, we used the ESRGAN model to increase the spatial resolution of each olive tree crown image in the OTC data set to obtain sharper edge contours, higher contrast, and more realistic and natural textures (Figure 3), which helped enable the deep learning model to obtain the edge information of the tree crown more accurately and preserve the useful hidden features in the image to the greatest extent.



Figure 3. Super-resolution (SR) reconstruction effect.

2.4. Data Augmentation

To obtain the best accuracy, deep learning models often require many images as a data set for training. However, the amount of images sampled in the study area is frequently limited, which would lead to overfitting during model training [41]. Therefore, a data augmentation method was used to address the problem in this study, which would perform a series of transformations on the images to expand the number of samples in the data set and enhance the model's generalization ability. In this paper, the original UAV images for the study area were geometrically transformed by rotating, flipping, and zooming (Figure 4). The geometrically transformed data allowed the model to better acquire olive crown features of different angles, sizes, and shapes as well as improved adaptability under various conditions.

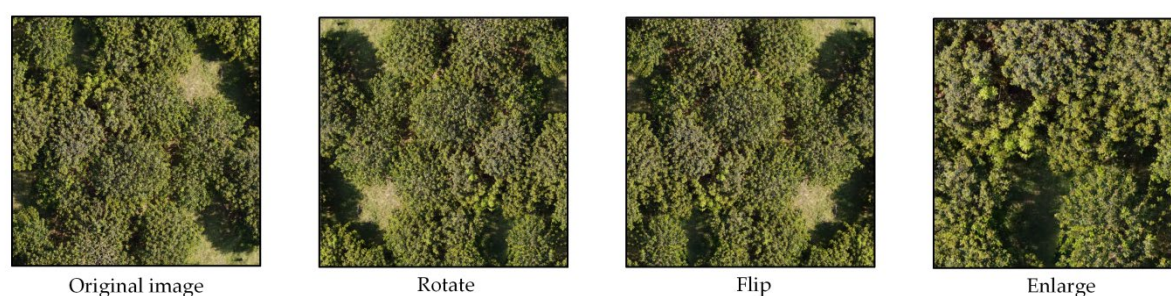


Figure 4. Geometric transformations.

The light conditions of olive canopies vary in different regions of the study area due to the topography, leading to differences in their brightness (value) in the images. Moreover, the study area contains olive trees at different growth stages, and the saturation (saturation) and hue (hue) of olive canopies fluctuate at various stages of development due to the influence of chlorophyll and other factors. To address this issue, we transformed the color space of the original OTC data set from the RGB model to the HSV model. As shown in Figure 5a, hue has a relatively smaller range, mainly within 82.28–86.79, whereas saturation and brightness have a larger ranges, mainly within 100.39–121.36 and 92.50–107.93, respectively.

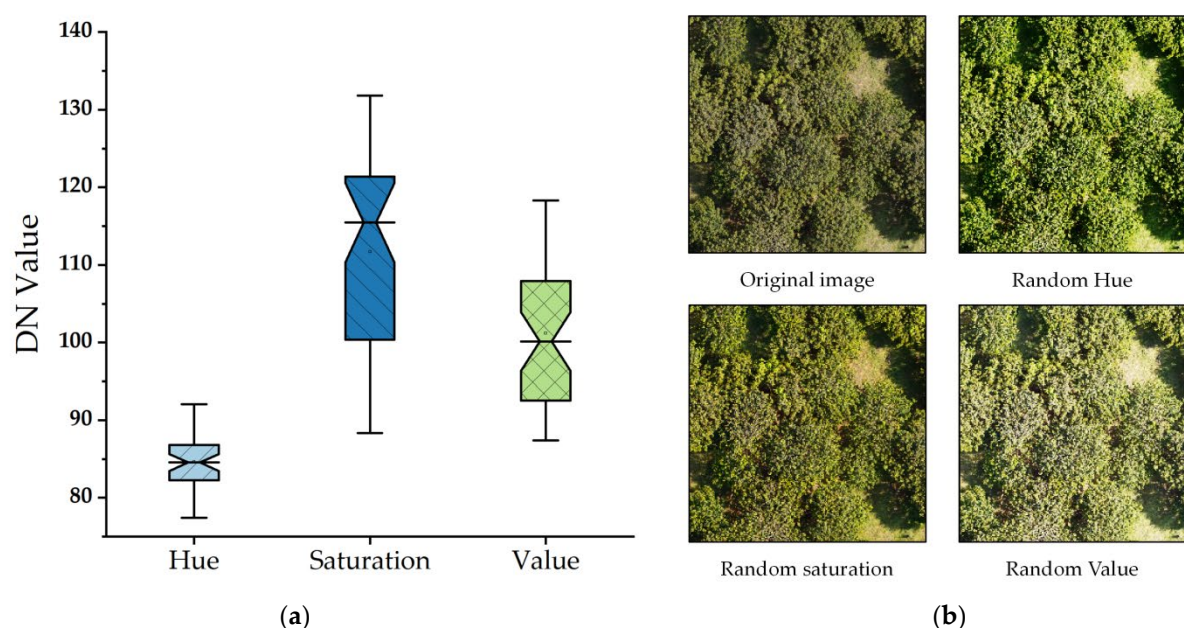


Figure 5. Spectral augmentation: (a) HSV boxplot; (b) spectral transformation.

2.5. Selection of Test Subareas and Measurement of Crown

In this study, four parts of the study area (i.e., A–D) with distinct and representative crown distribution characteristics were selected as typical experimental regions (Figure 6). The olive trees in part A were unevenly distributed and had some differences in crown size; in part B, trees were less spaced and more densely distributed, but there was no obvious overlap, and the degree of differentiation was obvious; in part C, the olive trees were densely spaced and evenly distributed with a large crown size and a high degree of overlap and adhesion; in part D, the olive trees were evenly distributed with no obvious overlap or adhesion and had a relatively uniform crown size. Since measuring the actual crown area in the field is challenging due to the large and irregular crown and great tree height of olive, we derived the actual crown area by manually vectorizing the olive crown based on UAV images using the ArcGIS 10.2 software. In the four experimental subareas, a detailed outline of all the olive canopies was first drawn and then the area of each crown was calculated using the Calculate Geometry tool in ArcGIS 10.2.

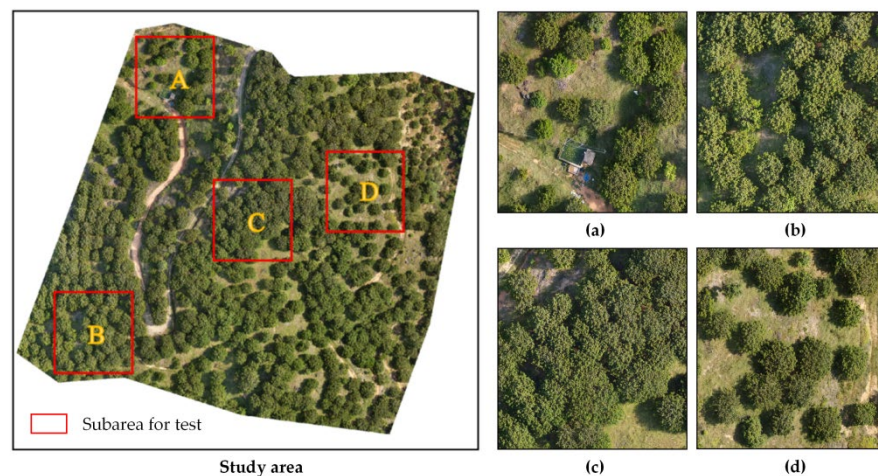


Figure 6. Location of the four test subareas (A–D): (a–d) details of olive trees in the A–D test subareas, respectively.

2.6. U²-Net Deep Learning Model

Achieving accurate crown detection requires integration of high-resolution local information and low-resolution global information, that is, utilizing multiple scale features that can produce better results. Most of the current multi-scale feature extraction methods try to extract local and global information from the features obtained through the backbone network by designing new modules to better utilize the existing image classification backbone. To better extract local and global features, Qin et al. proposed a novel deep learning model, U²-Net, and its structure is shown in Figure 7. Instead of adding more complex modules and strategies, the U²-Net model adopts a novel and simple architecture that directly extracts multi-scale features by stages, enabling accurate tree crown contours detection.

The model's overall design framework is a two-level nested U-shaped network structure. Each level follows an encoder–decoder structure similar to that in the U-Net model. The outermost layer of the U²-Net model is eleven separate large U-shaped structures composed of encoder and decoder blocks, each of which is filled with small Residual U-blocks (RSUs). In the first four encoding stages (En_1~4), more local and global features are obtained by increasing the number of convolutional layers to expand the receptive fields. Among them, the resolution of the feature maps in the top encoders EN_5 and EN_6 is relatively low, and further downsampling of these feature maps will result in the loss of useful feature information. Therefore, in the EN_5 and EN_6 stages, we replaced the pooling and upsampling operations with *dilated convolution*, which means that the feature

maps of EN_5 and EN_6 had the same high resolution as their input feature maps. The decoding phase was similar to the encoding phase, in which stepwise upsampling, merging, and convolution were employed to encode the high-resolution feature map, reducing the loss of details caused by direct upsampling. In addition, we also added a feature map fusion module of different scales to generate a saliency probability map. The principle was as follows: an image with a resolution of 288×288 was taken as an example, U²-Net first passed through a 3×3 convolutional layer and a sigmoid function to generate six feature probability maps of different sizes from EN_6, DE_5, DE_4, DE_3, DE_2, and DE_1: $S^{(6)}_{side}$ ($1 \times 9 \times 9$), $S^{(5)}_{side}$ ($1 \times 18 \times 18$), $S^{(4)}_{side}$ ($1 \times 36 \times 36$), $S^{(3)}_{side}$ ($1 \times 72 \times 72$), $S^{(2)}_{side}$ ($1 \times 144 \times 144$), and $S^{(1)}_{side}$ ($1 \times 288 \times 288$). Then, the feature probability maps of different sizes were upsampled with the sampling ratios of 32, 16, 8, 4, 2, and 1, respectively, to obtain six $1 \times 288 \times 288$ feature maps, which were fused using a cascade operation of 1×1 convolutional layer and a Sigmoid function. Finally, the tree crown prediction map (288×288) was generated with the smallest error. The new design enabled the network to extract multi-scale features and increase the depth of the whole model architecture without reducing the feature mapping resolution or significantly increasing the memory and computing cost.

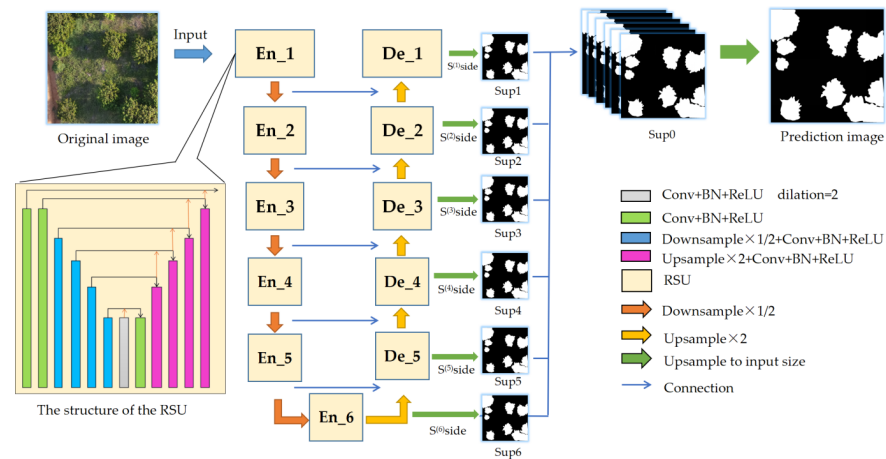


Figure 7. Structure of the U²-Net model. En represents the encoder, De represents the decoder, $S^{(n)}_{side}$ represents the side output saliency maps, and Sup represents the upsampling ratio.

2.6.1. RSU Structure

In the task of detection and segmentation of the image targets, local and global contextual information is a crucial factor affecting accuracy. Among the convolutional neural networks (CNNs) that have emerged in recent years, small convolutional filters (size of 1×1 or 3×3) are widely used for feature extraction with their compact size, such as VGG, ResNet, and DenseNet. These small convolutional filters are very popular in computer vision tasks due to the fact of their smaller storage footprint and good computational efficiency during operation. However, because the receptive field of the 1×1 or 3×3 convolutional filter is too small to obtain global context information, the output feature map of the shallow layer contains only local features. The most direct is to enlarge the receptive field to achieve more global information at high-resolution feature maps from shallow layers. Referenced from the classical U-Net network, a novel Residual U-block (RSU) was designed to capture multi-scale features at the intra-stage. The structure of the RSU (C_{in} , M , and C_{out}) is shown in Figure 8, where C_{in} and C_{out} denote the number of feature map input and output channels, respectively, while M denotes the number of channels in the internal layers of the RSU and L is the number of layers contained in the encoder. The RSU consists of three main parts:

- (1) The convolution layer located at the outermost layer first transforms the input feature map $x(H, W, \text{and } C_{in})$ into an intermediate map $F1(x)$ with an output channel of C_{out} . The role of this convolution layer is for extracting local features of the image;
- (2) The transformed intermediate map $F1(x)$ is input into a u-structure symmetric encoder–decoder with a height L . This structure we show it with $U(F1(x))$ that can extract multi-scale information and reduce the loss of contextual information caused by upsampling. When L is larger, the deeper the number of layers the RSU has, the more pooling operations are performed, the larger range of receptive fields, and the richer the local and global contextual features obtained;
- (3) The local features and the multi-scale features are fused by a residual connection: $H_{RSU}(x) = U(F1(x)) + F1(x)$.

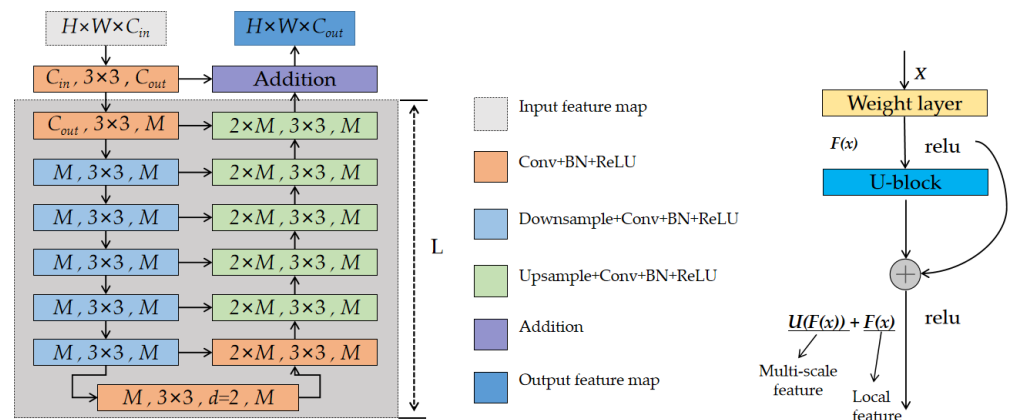


Figure 8. Structure of the residual U-block.

2.6.2. Loss

In the training process, the model output contained not only the final crown prediction map but also the feature maps of the previous six different scales ($S^{(1)}_{side} \sim S^{(6)}_{side}$). Therefore, in the process of training the model, we not only supervised the final output result map of the network but also supervised the feature maps of different scales in the middle; therefore, this paper used a deep supervision method similar to holistically nested edge detection (HED, Equation (1)) [42] and outputted seven losses per iteration. These losses were used to tune the model parameters to delineate overlapping tree crowns with minimal error.

$$L = \sum_{m=1}^M w_{side}^{(m)} l_{side}^{(m)} + w_{fuse} l_{fuse} \quad (1)$$

where $l_{side}^{(m)}$ ($M = 6$, as Sup1~6 in Figure 7) is the loss of the side output saliency map $S_{side}^{(m)}$, and l_{fuse} is the loss of the final fusion output saliency map S_{fuse} . $w_{side}^{(m)}$ and w_{fuse} are the weights of each loss term. For each term l , we used the standard binary cross-entropy to calculate the loss as shown in Equation (2) below:

$$l = - \sum_{(i,j)}^{(H,W)} [P_{G(i,j)} \log P_{S(i,j)} + (1 - P_{G(i,j)}) \log (1 - P_{S(i,j)})] \quad (2)$$

where (i,j) is the pixel coordinates and (H, W) is the image size, height and width, respectively. $P_{G(i,j)}$ and $P_{S(i,j)}$ denote the pixel values of the ground truth and the predicted saliency probability map, respectively. The training process tries to minimize the overall loss L (Equation (1)). We chose the fusion output l_{fuse} as our final saliency map in the testing process.

2.7. Accuracy Assessment

The intersection over union (IoU), is frequently used as a precision metric for semantic segmentation tasks. In the reference (R) and the prediction (P) mask, the *IoU* is indicated by a ratio of the number of pixels in both masks to the total number of pixels. By comparing the real category of samples with the model's predicted results, it can be classified into the following four cases: true positive (TP), where the predicted value of olive crown is consistent with the real value; false positive (FP), where the actual situation is the background but is incorrectly predicted as a crown; false negative (FN), where the crown in the real scene is not correctly identified; true negative (TN), where the background is consistent with the real value. In addition, precision (Equation (4)) and recall (Equation (5)), overall accuracy (OA, Equation (6)), and *F1-Score* (Equation (7)) are used as evaluation metrics to assess the model. The higher the precision, recall, OA, and *F1-Score*, the closer the predicted value is to the true value.

$$IoU = \frac{|R \cap P|}{|R \cup P|} \quad (3)$$

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{TP + FN} \quad (5)$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

$$F1-Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (7)$$

The coefficient of determination (R^2), mean relative error (MRE, Equation (8)), and root mean square error (RMSE, Equation (9)) were chosen to test the reliability of the U²-Net model for crown area extraction, where an R^2 closer to 1, indicates a higher correlation between the predicted and the measured values, and smaller values of MRE and RMSE suggest that the predicted values are closer to the measured values. The formulas are listed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (8)$$

$$MRE = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \times 100\% \quad (9)$$

3. Results

The hardware and software parameters in this study are shown in Table 1. One of the major advantages of the U²-Net model is that it does not require a pretrained model based on large data sets. Therefore, we used the Xavier [43] method to initialize all the convolutional layers in the model and used the Adam optimization algorithm [44] to adjust the model training process. The model batch size was set to 14, the number of iterations (epoch) was set to 300, the loss weights were set to 1, and the rest of the hyperparameters were set to the default values (initial learning rate = 0.003, betas = (0.9, 0.999), eps = 1×10^{-8} , and weight decay = 0). We configured the model according to the given parameters and input the OTC image data set for iterative training, which required a total of 18 h of training time.

Table 1. Parameters of the software and hardware.

Items	Parameters and Versions
CPU	Intel® Core™ i7-8700 @3.20 GHz
RAM	32 GB DDR4 2666 MHz
SSD	HS-SSD-C2000Pro 2 TB
GPU	Nvidia GeForce RTX 3060 (12 GB)
OS	Windows 10 Professional (DirectX 12)
ENVS	PyTorch 1.9.0 + Python 3.8

3.1. ESRGAN Model Performance Deviation

To evaluate the effect of ESRGAN-based enhancement processing on the model's performance, we compared the tree crown recognition results based on low-resolution (LR)-based images and ESRGAN-enhanced super-resolution (SR) images. The results (Figure 9) showed that after using ESRGAN enhancement, each accuracy index of the model increased to varying degrees, among which the precision increased by 9.45%, the recall rate increased by 5.68%, the F1-score increased by 7.63, the OA increased by 5.72, and the IoU increased by 6.4%. Therefore, the ESRGAN model significantly improve the crown prediction accuracy.

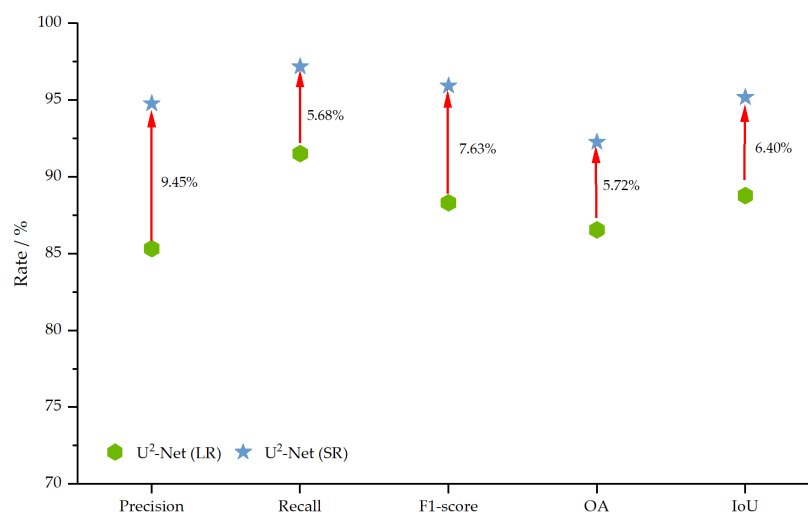


Figure 9. Comparison diagram of the performance of different resolution models.

3.2. Evaluation of Tree Crown Number Extraction Results

The results of olive crown extraction using the U²-Net model in four subareas are shown in Figure 10. The figure suggests that all test subareas demonstrated good extraction results. The density of olive tree crown in subareas A and D was lower; therefore, the trees were well distinguished from each other, and the overall outline of the crown was clear and the shape was consistent with the actual one. The density of the olive trees in subarea B was higher than that in subareas A and D. Nonetheless, there was a certain gap between most of the crowns in this area so that the U²-Net model could also distinguish single tree crowns well, with only a few crowns adhering to each other. On the contrary, the tree density in subarea C was higher, leading to the crowns overlapping each other more; therefore, the distinguished result in this subarea was not as good as that in other subareas. According to the prediction results of different subareas, it is clear that depression degree and tree spacing both had the greatest influence on the U²-Net model. Severe crown shading and closer tree spacing both make our model unable to distinguish each tree well and the “sticky” phenomenon occurs. In the four subareas, low vegetation under the forest was incorrectly identified as olive trees due to the similar spectral characteristics, and the crown was missed due to the shading.

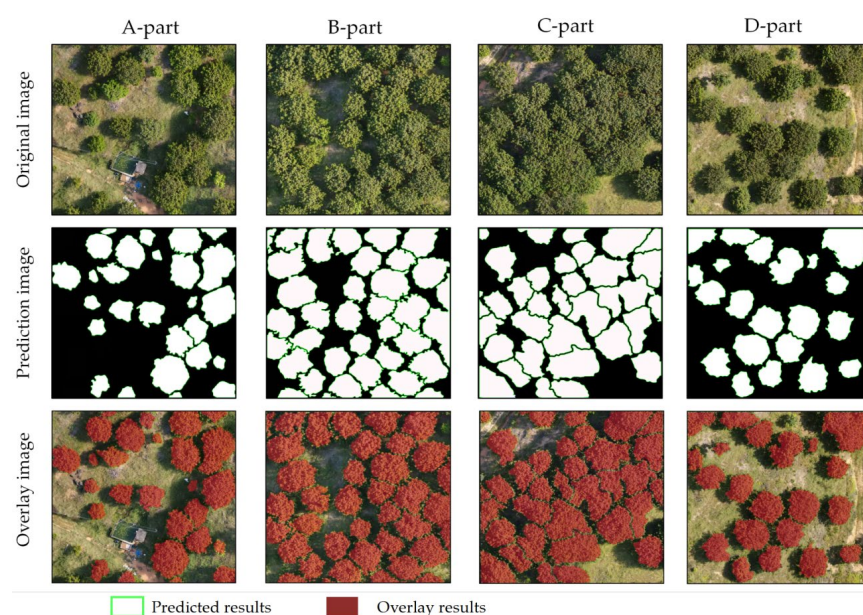


Figure 10. Segmentation results using the U²-Net model.

The quantitative evaluation results of the number of crowns extracted by the U²-Net model in different subareas are shown in Figure 11. The four subareas obtained a high accuracy rate with an F1-Score higher than 93.86% and overall accuracy exceeding 91.55%. By analyzing the prediction accuracy of olive trees in each subarea, it was found that U²-Net showed excellent performance in subareas A and D, where olive tree depression was low and crown edge adhesion and overlap were low, with overall accuracies of 97.41% and 98.38%, respectively, and F1-Scores of 96.49% and 98.08%. As the tree density increased, it made the adhesion and overlap of crown edges more serious, resulting in a slight decrease in the performance of the U²-Net model in subareas B and C, with an accuracy of only 94.01% and 90.60%, OA of 93.42% and 91.55%, and F1-Scores of 95.38% and 93.86%, respectively. It can be seen that the U²-Net model has good applicability in regions with low crown density. The statistics from the four subareas showed that the average values of IoU, OA, and F1-Score were 92.27%, 95.19%, and 95.95%, respectively, indicating that this method is feasible for olive tree crown extraction.

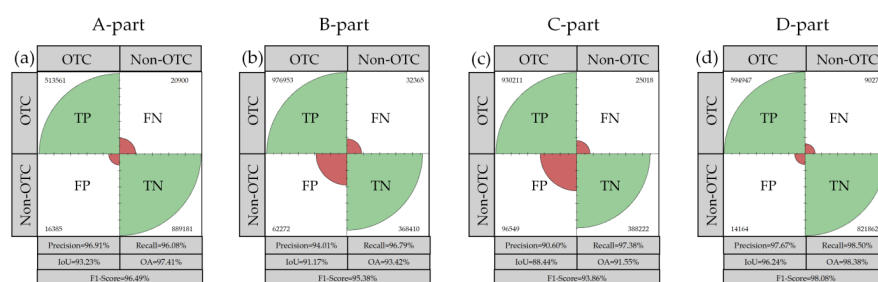


Figure 11. Extraction accuracy of olive tree crown number using the U²-Net model. (a–d) indicate the accuracy evaluation plots of four test subareas (A–D). In the figure, the green part indicates the proportion of pixels correctly classified as OTC and non-OTC, and the red part indicates the proportion of pixels incorrectly classified as OTC and non-OTC.

3.3. Accuracy Evaluation of Different Deep Learning Models

To investigate the applicability of the U²-Net model, this study compared its image segmentation and extraction results with three other mainstream deep learning models (i.e., HRNet, U-Net, and DeepLabv3+). DeepLabv3+ and U-Net use PascalVOC pretraining weights, HRNet uses CITYSCAPES pretraining weights, and all four models used the

same OTC image data set for training. The test visualization extraction results are shown in Figure 12. It can be seen that the extraction effects of the four deep learning models in the same experimental area had no apparent misclassification phenomenon, indicating that these models can well distinguish the target features from the background. However, when observing the details, we can see that the overall omission of HRNet and U-Net is serious, and the completeness of the extracted crown was poor. Compared with the previous two models, the DeepLabv3+ model had a superior overall effect. The model can detect smaller olive crowns, but the overall omission phenomenon still exists. The U²-Net model used in this paper significantly reduces the sensation of misclassification and omission. Although there are still some defects, the effect of crown edge segmentation is excellent. Although there remain several canopies adhering to each other, the U²-Net model used in this study significantly reduces misclassification and omission and has good crown edge segmentation.

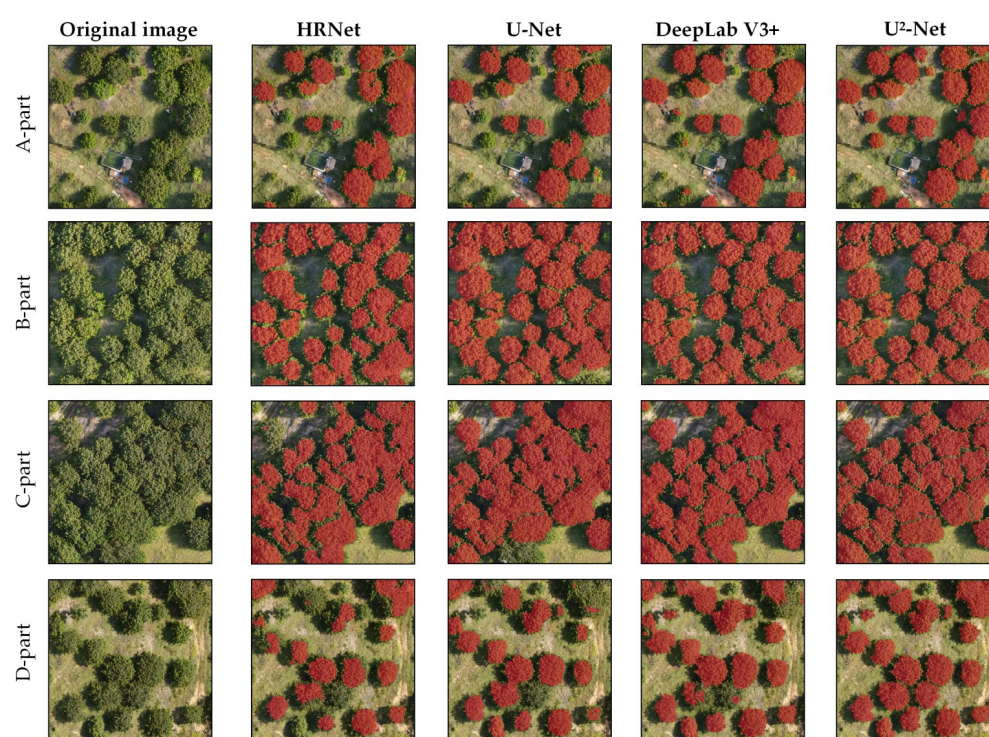


Figure 12. Segmentation results of olive tree crown for the different models. The red mask area in the figure represents the overlay prediction results for each model.

The results of the quantitative analysis of the extraction accuracy of the four models are shown in Figure 13. Although HRNet obtained the highest precision rate of 95.39%, the performance of other accuracy indicators was not as good as the other three models. DeepLabv3+, using ResNet-101 as the backbone network, outperformed U-Net in terms of overall accuracy. Although the precision of the U²-Net model with RSU as the backbone network was slightly lower than that of HRNet, it had obvious advantages in the recall, IoU, OA, and F1-Score. Among them, IoU, overall accuracy, and F1-Score increased by 14.03–23.97 percentage points, 7.57–12.85 percentage points, and 8.15–14.78 percentage points, respectively, compared with the other three models, indicating that the U²-Net model can achieve higher accuracy with a small amount of sample data, which can meet accuracy requirements for actual application scenarios.

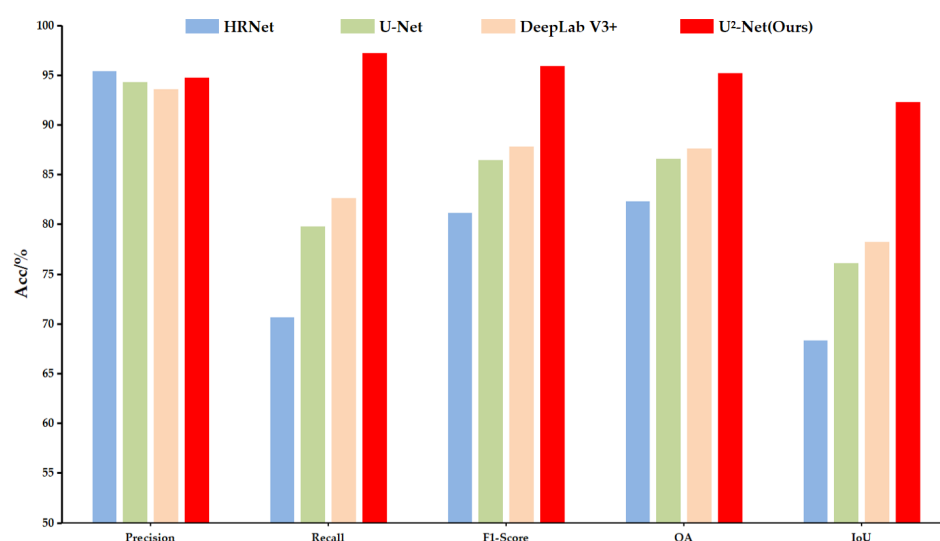


Figure 13. Extraction accuracy of olive tree crown number using different models.

3.4. Evaluation of Tree Crown Area Extraction Accuracy

The results of the quantitative evaluation of crown area extraction using the U²-Net model in different experimental areas are shown in Table 2. Subarea D, which has low crown density and good lighting conditions, achieved the highest accuracy with an RMSE and MRE of 3.01% and 11.87%, respectively. In contrast, the extraction accuracy of subarea A, which has a similar olive tree distribution to subarea D, had a lower extraction accuracy. The explanation for this is that despite the low crown density of subarea A, there were significant differences in the crown shape and height of olive trees in this area, resulting in the shadows cast by the tall olive trees blocking the surrounding smaller olive trees. Thus, only the top canopies with better lighting conditions can be recognized well, resulting in a smaller predicted crown area in subarea A. Despite the high crown density in subarea B, there was a certain gap among individual trees with similar tree height; thus, the shadow of nearby trees was less affected, resulting in a predicted value close to the measured value. Therefore, the overall accuracy of subarea B was higher, with an RMSE and MRE of 3.85% and 12.27%, respectively. The worst accuracy was in area C and its RMSE and MRE were the lowest among all subareas, with only 6.72% and 16.38%, respectively.

Table 2. Extraction accuracy of crown area using U²-Net model.

Subarea	Number of Trees	RMSE	MRE/%
A	25	3.95	16.37
B	38	3.85	12.27
C	34	6.72	16.38
D	20	3.01	11.87
Mean	29	4.78	14.27

Comparing the relationship between the predicted and measured crown areas in the four experimental subareas (Figure 14), it can be seen that the R^2 of the three experimental subareas, A, B, and D, were greater than 0.95. Most points were distributed around the 1:1 straight line, indicating that the extracted area by the U²-Net model in these areas was highly consistent with the measured values. However, the coefficient of determination, R^2 , in subarea C was relatively low. Due to the severe overlap and adhesion of the olive trees in this area, most of the extracted areas were smaller than the measured values as shown by the RMSE and MRE. On the other hand, the extracted area values of different experimental subareas were generally close to the measured values. Therefore, the U²-Net

model had high accuracy in extracting the crown area of olive trees and could meet the demand of improving the accuracy of a crown area in forestry surveys.

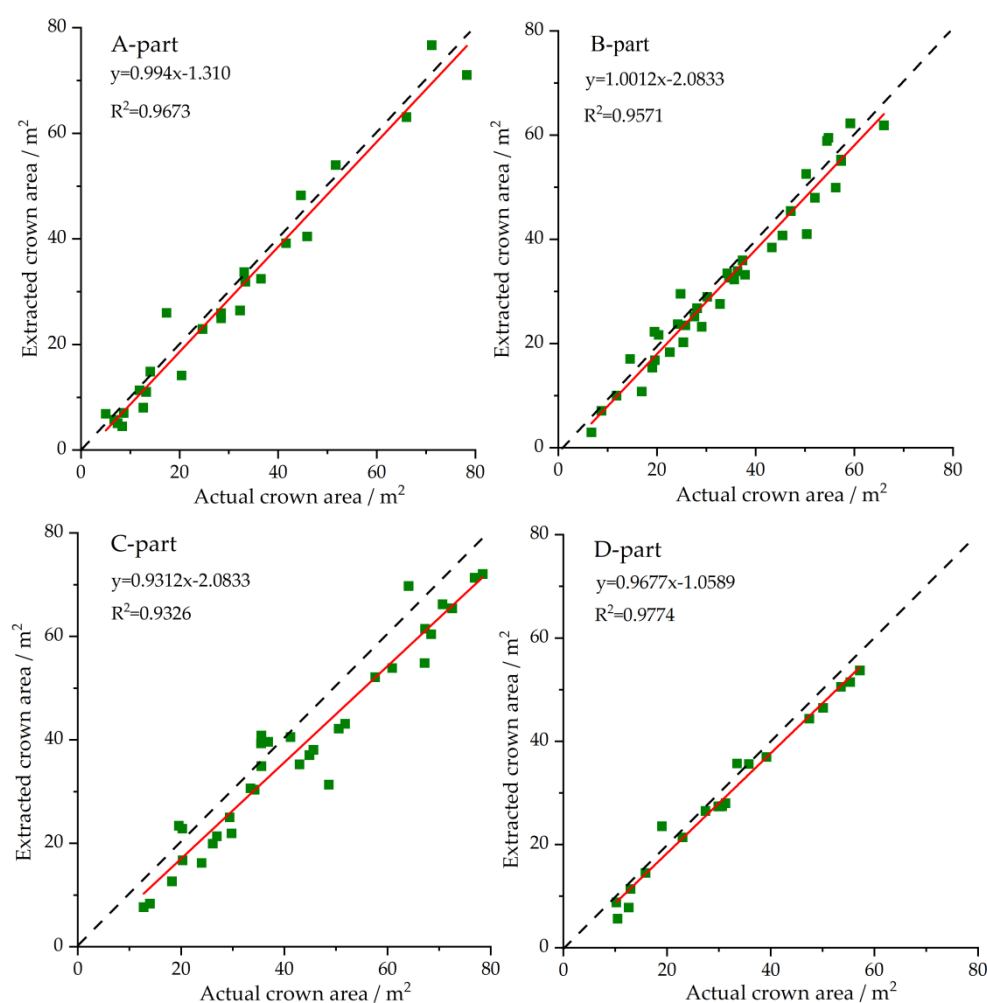


Figure 14. Relationship between extracted area and measured area using the U²-Net model.

The errors of the four models in the olive crown extraction task are shown in Figure 15. From this, we can see that the U²-Net model had the lowest error rates both in RMSE (4.78) and MRE (14.27%), indicating that its extracted area values were closest to the measured value. The DeepLab V3+ model performed similarly to U²-Net with an RMSE value of 5.42 and an MRE value of 16.31%, while U-Net and HRNet performed worse with RMSEs of 7.88 and 9.24, respectively, and MREs of 21.36% and 24.59%, respectively. Therefore, U²-Net performed better than the other three models.

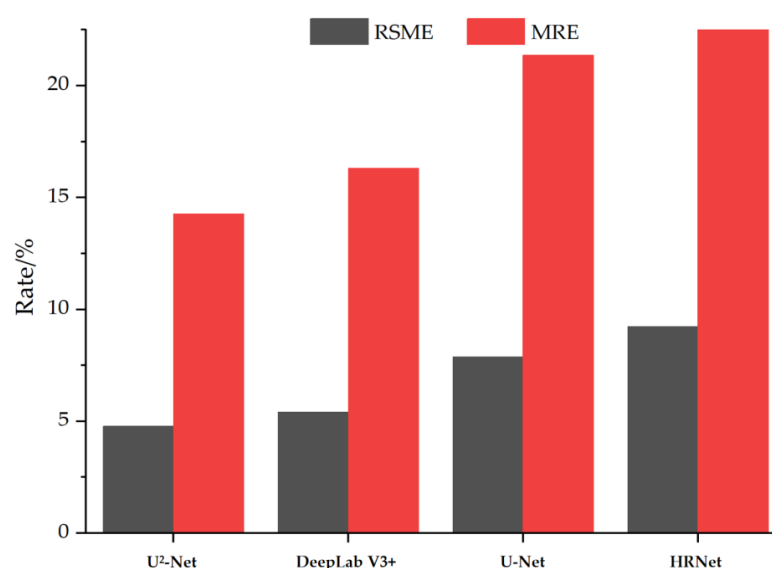


Figure 15. Comparison of the error rates for the different models.

4. Discussion

This paper proposes a method that combines a consumer-grade UAV remote sensing platform and a deep learning U²-Net model to achieve automatic extraction and area parameter prediction of olive tree crown at the single-tree level. The results show that the method proposed in this study had high accuracy for crown segmentation and area prediction that can meet the needs of olive crown information extraction in agricultural production. Compared with the traditional measurement methods (i.e., manual way and satellite remote sensing), the UAV remote sensing monitoring method used in this study has great advantages. Traditional manual measurements are time-consuming, labor-intensive, error-prone, and difficult to monitor in real time. Satellite remote sensing platforms are difficult to ensure accuracy and time continuity, and its high-resolution data requires high costs. However, the UAV remote sensing platform has high temporal and spatial resolution, flexible operation, and low cost, which could be an essential supplement to conventional monitoring methods for agricultural and forestry resources. In addition, with the wide application of UAVs, the amount of remote sensing data has increased dramatically, and the automation and intelligence of data processing require more effective algorithms. Machine learning and deep learning are important algorithms to solve the current predicament. However, due to the limited analysis ability of machine learning, it is difficult to extract features with highly complex, deep, and nonlinear relationships in the target image. Therefore, this paper adopted deep learning technology developed in recent years that can extract complex nonlinear features from massive high-dimensional data, and it has higher accuracy and better generalization and stability. Therefore, the combination of UAV and deep learning proposed by our study will be the development trend of intelligent processing of remote sensing data and automatic extraction of information in the future, providing technical support for the scientific management of agricultural and forestry resources.

We also trained and evaluated current mainstream three models (i.e., HRNet, U-Net, and DeepLabv3+). The results show that compared to the other models (the original U-Net, HRNet, and DeepLabv3+), the U²-Net model performed better in accurately segmenting tree crowns in complex backgrounds. Due to the well-designed architecture of U²-Net for extracting and integrating high-resolution local information and low-resolution global information, it is more accurate and has a lower prediction error. The RMSE and MRE of the U²-Net model were 4.78 and 14.27%, respectively. Compared with the three other models, the IoU, overall accuracy, and F1-Score of the U²-Net model increased by 14.03–23.97 percentage points, 7.57–12.85 percentage points, and 8.15–14.78 percentage

points, respectively. At the same time, the U²-Net model had certain similarities with DeepLabv3+. To integrate multi-scale information, they introduced the encoder–decoder framework commonly used in semantic segmentation and balanced the accuracy and its time-consuming nature by using dilated convolution. The difference is that DeepLabv3+ uses multi-scale atrous convolution cascades to capture multi-scale feature information, while U²-Net uses a nested architecture. Therefore, U²-Net with nested architecture had better performance than DeepLabv3+ with cascaded architecture, where the precision, recall, F1-score, OA, and IoU increased by 1.14 percentage points, 14.54 percentage points, 8.15 percentage points, 7.57 percentage points, and 14.03 percentage points, respectively, which shows that our proposed method had obvious superiority.

However, it is important to note that the proposed method still has some limitations and is open to improvement:

- (1) Although the proposed model was shown to be accurate and efficient, it does have limitations. By analyzing the extraction results from different experimental subareas (i.e., A–D), it can be found that even if aerial photography is conducted at noon, the UAV images still cannot completely avoid the problem of shadows induced by light. Due to the shadows, it is difficult for the model to extract the olive tree crowns that are obscured by them, resulting in lower predicted values than the measured values. Furthermore, because the UAV images used provide limited spectral information due to the limited bands (only visible light), the phenomenon of “different objects sharing similar spectrum” was obvious. Some other low vegetation often has similar spectral characteristics to a tree crown, which is easily misjudged by the model as a tree crown, thus affecting the accuracy of crown area extraction. Moreover, there are some large olive trees in the study area, which may have several crowns and multiple tree vertices, resulting in recognition error as well. Since the model cannot tell whether the adjacent crowns belong to the same tree, one tree with several subcrowns is often over-segmented and identified as multiple trees. Therefore, more studies are needed to explore the introduction of a crown height model (CHM) with elevation information and image data with visible vegetation indices to increase the recognition accuracy.
- (2) When using UAV imagery for crown extraction, factors, such as topography, vegetation type, and crown density of the study area, significantly affect the extraction accuracy. This study achieved good crown extraction results using the U²-Net model mainly due to the relatively flat terrain, regular distribution of olive trees, homogeneous tree species, and low overall crown density, which is similar to the experimental results conducted by Kuikel et al. [45] using banana tree planting areas. However, other studies have demonstrated that when the terrain has a specific slope, the crown parameter extraction accuracy will dramatically decrease. Casapia et al. [46] acknowledged this theory in their work as well: crown extraction can be hampered by complex vegetation types, densely overlapping canopies, and smaller tree spacing. Therefore, future studies should focus on achieving high-precision crown extraction under more complex environmental conditions.

Future research in this topic should take the following directions: first, the deployment of deep learning models in the embedded systems of UAVs for real-time prediction of growth characteristics, health conditions, crop yields, and other applications in precision agriculture. Second, in order to attain high accuracy in deep learning, a large amount of training data is required. As a result, additional image data from other growing areas are required to enrich the OTC data set. Finally, the current model is only applicable to olive trees, and future research could look towards developing a model that works for many kinds of economically significant forests.

5. Conclusions

In our study, we used a consumer-grade UAV equipped with a visible-light sensor to acquire orthophotos from an olive tree planting area, constructed an OTC image data set, and conducted an automated extraction study of olive tree crown information using a novel model of deep learning called U²-Net. The results of olive tree crown extraction revealed that the U²-Net model achieved good detection and segmentation results in different experimental subareas. The highest recognition accuracy was achieved in experimental subarea D with relatively large olive tree spacing and a small difference in crown width, with a precision and recall of 97.69% and 98.55%, respectively, indicating that the model has good applicability in forest stands with low crown density. Compared with three mainstream deep learning models (i.e., HRNet, U-Net, and DeepLabv3+) in crown number extraction, the U²-Net model had a higher cross-merge ratio, overall accuracy, and F1-Score, which improved by 14.03~23.97%, 7.57~12.85%, and 8.15~14.78%, respectively. In addition, the U²-Net model also performed better in the crown area, the model's predicted area was closer to the measured area, and the average relative error and root mean square error of the four experimental areas were low at 4.78 and 14.27, respectively, while R^2 was higher than 0.93. In summary, the method for extracting tree crown and crown area estimation using UAV visible images combined with the U²-Net deep learning model can be employed effectively. Based on the UAV platform, this method can acquire and analyze a large amount of high-precision olive crown information at a low cost, providing accurate data support for higher-level applications, allowing for the development of many potential applications and more accurate monitoring and management in the forestry field.

Author Contributions: Conceptualization, Z.Y. and H.Z.; methodology, Z.Y. and H.Z.; software, Z.Y. and H.Z.; validation, H.Z., Q.G., J.W., and J.Z.; formal analysis, H.Z. and Y.L.; investigation, Z.Y. and H.Z.; resources, H.Z.; data curation, Z.Y.; writing—original draft preparation, Z.Y. and H.Z.; writing—review and editing, H.Z., Q.G., and J.Z.; funding acquisition, H.Z., H.D., and K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (31901298), Natural Science Foundation of Fujian Province (2021J01059, 2020J05021), and the Key R&D plan of the Department of Tibet Autonomous Region Science and Development (XZ202001ZY0056G).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the developers in the GitHub community for their open-source U²-Net deep learning projects.

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

References

1. Arampatzis, G.; Hatzigiannakis, E.; Pisinaras, V.; Kourgialas, N.; Psarras, G.; Kinigopoulou, V.; Panagopoulos, A.; Koubouris, G. Soil water content and olive tree yield responses to soil management, irrigation, and precipitation in a hilly Mediterranean area. *J. Water Clim. Chang.* **2018**, *9*, 672–678. <https://doi.org/10.2166/wcc.2018.224>.
2. Montealegre, C.; Esteve, C.; Garcia, M.C.; Garcia-Ruiz, C.; Marina, M.L. Proteins in olive fruit and oil. *Crit. Rev. Food Sci. Nutr.* **2014**, *54*, 611–624. <https://doi.org/10.1080/10408398.2011.598639>.
3. Carletto, C.; Jolliffe, D.; Banerjee, R. From tragedy to renaissance: Improving agricultural data for better policies. *J. Dev. Stud.* **2015**, *51*, 133–148. <https://doi.org/10.1080/00220388.2014.968140>.
4. Xiong, J.; Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Poehnelt, J.; Congalton, R.G.; Yadav, K.; Thau, D. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS-J. Photogramm. Remote Sens.* **2017**, *126*, 225–244. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>.
5. Bokalo, M.; Stadt, K.J.; Comeau, P.G.; Titus, S.J. The Validation of the Mixedwood Growth Model (MGM) for Use in Forest Management Decision Making. *Forests* **2013**, *4*, 1. <https://doi.org/10.3390/f4010001>.
6. Li, Y.; Wang, W.; Zeng, W.S.; Wang, J.J.; Meng, J.H. Development of Crown Ratio and Height to Crown Base Models for Masson Pine in Southern China. *Forests* **2020**, *11*, 17. <https://doi.org/10.3390/f11111216>.

7. Narvaez, F.Y.; Reina, G.; Torres-Torriti, M.; Kantor, G.; Cheein, F.A. A Survey of Ranging and Imaging Techniques for Precision Agriculture Phenotyping. *IEEE-ASME Trans. Mechatron.* **2017**, *22*, 2428–2439. <https://doi.org/10.1109/tmech.2017.2760866>.
8. Gongal, A.; Amatya, S.; Karkee, M.; Zhang, Q.; Lewis, K. Sensors and systems for fruit detection and localization: A review. *Comput. Electron. Agric.* **2015**, *116*, 8–19. <https://doi.org/10.1016/j.compag.2015.05.021>.
9. Bargoti, S.; Underwood, J.P. Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *J. Field Robot.* **2017**, *34*, 1039–1060. <https://doi.org/10.1002/rob.21699>.
10. Aubry-Kientz, M.; Laybros, A.; Weinstein, B.; Ball, J.G.C.; Jackson, T.; Coomes, D.; Vincent, G. Multisensor Data Fusion for Improved Segmentation of Individual Tree Crowns in Dense Tropical Forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3927–3936. <https://doi.org/10.1109/jstars.2021.3069159>.
11. Bagheri, R.; Jouibary, S.S.; Erfanfard, Y. Canopy based aboveground biomass and carbon stock estimation of wild pistachio trees in arid woodlands using Geoeye-1 images. *J. Agric. Sci. Technol.* **2021**, *23*, 107–123.
12. Cho, M.A.; Malahlela, O.; Ramoelo, A. Assessing the utility WorldView-2 imagery for tree species mapping in South African subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 349–357. <https://doi.org/10.1016/j.jag.2015.01.015>.
13. Ferreira, M.P.; Wagner, F.H.; Aragao, L.; Shimabukuro, Y.E.; de Souza, C.R. Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS-J. Photogramm. Remote Sens.* **2019**, *149*, 119–131. <https://doi.org/10.1016/j.isprsjprs.2019.01.019>.
14. Wu, S.; Wang, J.; Yan, Z.; Song, G.; Chen, Y.; Ma, Q.; Deng, M.; Wu, Y.; Zhao, Y.; Guo, Z.; et al. Monitoring tree-crown scale autumn leaf phenology in a temperate forest with an integration of PlanetScope and drone remote sensing observations. *ISPRS-J. Photogramm. Remote Sens.* **2021**, *171*, 36–48. <https://doi.org/10.1016/j.isprsjprs.2020.10.017>.
15. Yan, S.; Jing, L.; Wang, H. A new individual tree species recognition method based on a convolutional neural network and high-spatial resolution remote sensing imagery. *Remote Sens.* **2021**, *13*, 479. <https://doi.org/10.3390/rs13030479>.
16. Gong, C.; Li, L.; Hu, Y.; Wang, X.; He, Z.; Wang, X. Urban river water quality monitoring with unmanned plane hyperspectral remote sensing data. In Proceedings of the 7th Symposium on Novel Photoelectronic Detection Technology and Applications, Kunming, China, 5–7 November 2020.
17. Gumma, M.K.; Kadiyala, M.D.M.; Panjala, P.; Ray, S.S.; Akuraju, V.R.; Dubey, S.; Smith, A.P.; Das, R.; Whitbread, A.M. Assimilation of remote sensing data into crop growth model for yield estimation: A case study from India. *J. Indian Soc. Remote Sens.* **2021**. <https://doi.org/10.1007/s12524-021-01341-6>.
18. Gale, M.G.; Cary, G.J.; Van Dijk, A.I.J.M.; Yebra, M. Forest fire fuel through the lens of remote sensing: Review of approaches, challenges and future directions in the remote sensing of biotic determinants of fire behaviour. *Remote Sens. Environ.* **2021**, *255*, 112282. <https://doi.org/10.1016/j.rse.2020.112282>.
19. Guimaraes, N.; Padua, L.; Marques, P.; Silva, N.; Peres, E.; Sousa, J.J. Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities. *Remote Sens.* **2020**, *12*, 35. <https://doi.org/10.3390/rs12061046>.
20. He, H.Q.; Yan, Y.; Chen, T.; Cheng, P.G. Tree height estimation of forest plantation in mountainous terrain from bare-earth points using a dog-coupled radial basis function neural network. *Remote Sens.* **2019**, *11*, 21. <https://doi.org/10.3390/rs11111271>.
21. Jurado, J.M.; Ortega, L.; Cubillas, J.J.; Feito, F.R. Multispectral mapping on 3d models and multi-temporal monitoring for individual characterization of olive trees. *Remote Sens.* **2020**, *12*, 1106. <https://doi.org/10.3390/rs12071106>.
22. Torresan, C.; Berton, A.; Carotenuto, F.; Di Gennaro, S.F.; Gioli, B.; Matese, A.; Miglietta, F.; Vagnoli, C.; Zaldei, A.; Wallace, L. Forestry applications of UAVs in Europe: A review. *Int. J. Remote Sens.* **2017**, *38*, 2427–2447. <https://doi.org/10.1080/01431161.2016.1252477>.
23. Egli, S.; Hoepke, M. CNN-Based Tree Species Classification Using High Resolution RGB Image Data from Automated UAV Observations. *Remote Sens.* **2020**, *12*, 3892. <https://doi.org/10.3390/rs12233892>.
24. Hao, Z.; Lin, L.; Post, C.J.; Mikhailova, E.A.; Li, M.; Yu, K.; Liu, J.; Chen, Y. 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. <https://doi.org/10.1016/j.isprsjprs.2021.06.003>.
25. Onishi, M.; Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **2021**, *11*, 903. <https://doi.org/10.1038/s41598-020-79653-9>.
26. 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. <https://doi.org/10.1016/j.compag.2020.105504>.
27. Dong, S.; Wang, P.; Abbas, K. A survey on deep learning and its applications. *Comput. Sci. Rev.* **2021**, *40*, 100379. <https://doi.org/10.1016/j.cosrev.2021.100379>.
28. Xie, S.; Yu, Z.; Lv, Z. Multi-disease prediction based on deep learning: A survey. *Cmes-Comput. Modeling Eng. Sci.* **2021**, *128*, 489–522. <https://doi.org/10.32604/cmes.2021.016728>.
29. Osco, L.P.; de Arruda, M.d.S.; Marcato Junior, J.; da Silva, N.B.; Marques Ramos, A.P.; Saito Moryia, E.A.; Imai, N.N.; Pereira, D.R.; Creste, J.E.; Matsubara, E.T.; 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. <https://doi.org/10.1016/j.isprsjprs.2019.12.010>.
30. Ferreira, M.P.; Almeida, D.R.A.d.; Papa, D.d.A.; Minervino, J.B.S.; Veras, H.F.P.; Formighieri, A.; Santos, C.A.N.; Ferreira, M.A.D.; Figueiredo, E.O.; Ferreira, E.J.L. Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. *For. Ecol. Manag.* **2020**, *475*, 118397. <https://doi.org/10.1016/j.foreco.2020.118397>.

31. Lou, X.W.; Huang, Y.X.; Fang, L.M.; Huang, S.Q.; Gao, H.L.; Yang, L.B.; Weng, Y.H.; Hung, I.K.U. Measuring loblolly pine crowns with drone imagery through deep learning. *J. For. Res.* **2022**, *33*, 227–238. <https://doi.org/10.1007/s11676-021-01328-6>.
32. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Huang, G.; Liu, Z.; Laurens, V.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
35. Xia, G.-S.; Yang, W.; Delon, J.; Gousseau, Y.; Sun, H.; Maitre, H. Structural high-resolution satellite image indexing. In Proceedings of the ISPRS Technical Commission VII Symposium—100 Years ISPRS—Advancing Remote Sensing Science, Vienna, Austria, 5–7 July 2010; pp. 298–303.
36. Xia, G.-S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. <https://doi.org/10.1109/tgrs.2017.2685945>.
37. Cheng, G.; Han, J.W.; Zhou, P.C.; Guo, L. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS-J. Photogramm. Remote Sens.* **2014**, *98*, 119–132. <https://doi.org/10.1016/j.isprsjprs.2014.10.002>.
38. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U-2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. <https://doi.org/10.1016/j.patcog.2020.107404>.
39. Wang, C.; Zhu, R.; Bai, Y.; Zhang, P.; Fan, H. Single-frame super-resolution for high resolution optical remote-sensing data products. *Int. J. Remote Sens.* **2021**, *42*, 8099–8123. <https://doi.org/10.1080/01431161.2021.1971790>.
40. Xiong, Y.; Guo, S.; Chen, J.; Deng, X.; Sun, L.; Zheng, X.; Xu, W. Improved SRGAN for Remote Sensing Image Super-Resolution across Locations and Sensors. *Remote Sens.* **2020**, *12*, 1263. <https://doi.org/10.3390/rs12081263>.
41. Su, D.; Kong, H.; Qiao, Y.; Sukkarieh, S. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Comput. Electron. Agric.* **2021**, *190*, 106418. <https://doi.org/10.1016/j.compag.2021.106418>.
42. Xie, S.; Tu, Z. Holistically-Nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1395–1403.
43. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
44. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
45. Kuikel, S.; Upadhyay, B.; Aryal, D.; Bista, S.; Awasthi, B.; Shrestha, S. Individual banana tree crown delineation using unmanned aerial vehicle (UAV) images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 581–585.
46. Tagle Casapia, X.; Falen, L.; Bartholomeus, H.; Cardenas, R.; Flores, G.; Herold, M.; Honorio Coronado, E.N.; Baker, T.R. Identifying and quantifying the abundance of economically important palms in tropical moist forest using uav imagery. *Remote Sens.* **2020**, *12*, 9. <https://doi.org/10.3390/rs12010009>.