Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Type of the Paper (Article)

An approach to detecting and mapping individual fruit trees integrated YOLOv5 with UAV remote sensing

Yongzhu Xiong [1,2,*], Xiaofeng Zeng [1,3*] , Jiawen Liao [1], Weiqian Lai [1,4], Yankui Chen[1,2] and Mingyong Zhu [1,2]

[1] School of Geography and Tourism, Jiaying University, Meizhou 514015 China; xiongyz@jyu.edu.cn (Y.X.); ljw1257871614@gmail.com (J.L.); 519958394@qq.com (Y.C.); 201701027@jyu.edu.cn (M.Z.);
[2] Guangdong Provincial Key Laboratory of Conservation and Precision Utilization of Characteristic Agricultural Resources in Mountainous Areas, Meizhou 514015, China;
[3] Guangzhou O.cn Network Technology Co., Ltd., Guangzhou 511400, China; zengxf12123@163.com;
[4] Satxspace (Dongguan) Technology Co., Ltd., Dongguan 523830, China; laiweigan@gmail.com;
* Correspondence: xiongyz@jyu.edu.cn (Y.X.) and zengxf12123@163.com (X.Z.); Tel.: 086-753-2186-956

Abstract: The location and number data of individual fruit trees are critical for planting area investigation, fruit yield prediction, and smart orchard management and planning. These data are conventionally obtained through manual investigation and statistics with time-consuming and laborious effort. Object detection models in deep learning used widely in computer vision could provide an opportunity for accurate detection of individual fruit trees, which is essential for rapidly obtaining the data and reducing human operations errors. This study proposes an approach to detecting individual fruit trees and mapping their spatial distribution by integrating deep learning with unmanned aerial vehicle (UAV) remote sensing. UAV remote sensing collected high-resolution true-color images of fruit trees in the experimental pomelo tree orchards in Meizhou city, South China. An image dataset of deep learning samples of individual pomelo trees (IPTs) was constructed through visual interpretation and field investigation based on the fruit tree images captured by UAV remote sensing. Four different scales of YOLOv5 (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) for object detection were selected to train, validate, and test on the image dataset of pomelo trees. The results show that the average precision (AP@0.5) of the four YOLOv5 models for validation reach 87.8%, 88.5%, 89.1%, and 90.7%, respectively. The larger the model scale, the higher the average accuracy of the detection result of validation. It suggests that YOLOv5x is a preferred high-accuracy model among the YOLOv5 family and is suitable to realize the detection of IPTs. The number of the IPTs in the study area was counted using YOLOv5x, and their spatial distribution map was made using the non-maximum suppression method and ArcGIS software. This study will provide primary data and technical support for smart orchard management in Meizhou city and other fruit-producing areas.

Keywords: individual fruit tree (IFT); individual pomelo tree (IPT) detection; deep learning; transfer learning; YOLOv5; remote sensing; unmanned aerial vehicle (UAV); spatial distribution

1. Introduction

Smart orchard management needs the data on the spatial and attributes of individual fruit trees (IFTs) in an orchard because these data play an important role in accurate planting area survey, disease and pest prevention and control, and fruit yield prediction. Traditionally, field investigations and statistics were used to collect these data including locations, spatial distribution, number of IFTs, and so on in orchards. These investigations are time-consuming, labor-extensive, and expensive, being however unable to meet the needs of smart orchard management. It is necessary to develop a fast, inexpensive, and accurate method for the investigation of IFTs to obtain these data.

Remote sensing images of fruit trees in a relatively large orchard can be captured by either satellite or aerial imaging. In the case of using satellite remote sensing, cloudy weather is a big challenge at first, making it difficult to detect fruit trees due to the possibly

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

poor quality of images captured. The limitation of spatial resolution of satellite images is another big challenge for the accurate detection of IFTs. Aerial imaging includes photography using a manned or unmanned aircraft. A manned aircraft is not suitable for the detection of IFTs due to the expensive costs and the inconvenient operations. Unmanned aerial vehicle (UAV) remote sensing is the best alternative to fulfill this task. Drones are a subset of UAVs and they are typically much small, lightweight, and cheap. A drone typically contains one or more very high-resolution cameras that can capture medium to high-quality images, depending on the altitude of flying [1]. UAV remote sensing has the advantages of automation, intelligence, and specialization to quickly obtain space remote sensing information such as land, resources, environment, and events, and conduct real-time processing, modeling, and analysis of advanced emerging aerial remote sensing technology solutions [2]. Recently, it is widely used in a lot of practice areas such as photogrammetry [3], precision agriculture [4], geohazard assessment [2], wildfire detection [5], and environmental monitoring [6, 7]. UAV remote sensing has great potential in acquiring the image data of fruit trees in orchards quickly and economically.

In the recent decade, with the progress of computer hardware equipment and the rapid development of artificial intelligence (AI) technology, convolutional neural networks (CNN) in deep learning, a core technology in AI, have pioneered new ways for object detection and feature extraction in remote sensing images [8, 9]. Many CNN architectures have been put forward for object detection in computer vision and image analysis, which can be divided into two categories, i.e., two-stage and one-stage models. Girshick *et al*. (2013) proposed a two-stage object detection model R-CNN (region-based convolutional neural network) based on classification problems [10]. Based on R-CNN, Fast R-CNN and Faster R-CNN were then proposed to improve the efficiency and accuracy. Redmon *et al*. (2016) proposed a single-stage based object detection model YOLO (you look only once) [11]. The YOLO model not only simplifies the size of the neural network but also improves the detection speed while improving the detection accuracy. Etten (2018) [12] proposed a pipeline (You Only Look Twice, or YOLT) based on YOLOv2 to fulfill rapid multi-scale object detection in large-scale satellite imagery. Yan *et al*. (2019) realized the recognition of *Rosa roxbunghii* in the natural environment based on improved Faster R-CNN with the average recognition accuracy of 11 classes of *Rosa roxbunghii* fruit to 92.01% [13]. Liu and Wang (2020) proposed a tomato disease and pest detection algorithm based on YOLO convolutional neural network [14]. Their results showed that YOLO outperforms Faster R-CNN. Xiong *et al*. (2020) proposed a visual detection method by using UAV images and YOLOv2 to detect green mangoes on the surface of the tree crown rapidly and estimate the number of mango fruits in orchards. Osco *et al*. (2021) present a comprehensive review of the fundamentals of deep learning applied in UAV-based imagery [8], providing a key reference to integrating deep learning with UAS remote sensing for the detection of IFTs.

More recently, several deep learning object detection and segmentation models have been progressively adopted in the detection and segmentation of individual trees such as olive, palm, and coconut trees based on high-resolution visible and LiDAR (light detection and ranging) images acquired from satellites and UAVs [1, 15-37]. Santos *et al*. (2019) [38] proposed and evaluated the usage of CNN-based methods combined with UAV high spatial resolution red-green-blue (RGB) imagery for the detection of law-protected tree species. Three state-of-the-art object detection methods were evaluated: Faster R-CNN, YOLOv3, and RetinaNet. RetinaNet achieved the most accurate results, having delivered 92.64% average precision. An analysis of satellite images by Brandt *et al*. (2020) has pinpointed individual tree canopies over a large area of West Africa [31]. Their results suggest that it will soon be possible, with certain limitations, to map the location and size of every tree worldwide [31, 32]. Safonova *et al*. (2021) [39] used Mask R-CNN and UAV images for olive tree crown and shadow segmentation to further estimate the biovolume of individual trees. Sun *et al*. (2022) counted the population number of trees in the subtropical megacity of Guangzhou and used an end-to-end tree-counting deep-learning framework (CMask R-CNN) in the regional-scale tree detection by delineating each tree crown [15].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Jintasuttisak *et al*. (2022) [1] carried out the automatic detection of crowded date palm trees in drone imagery using YOLOv5 and a comparison of one-stage object detection methods (YOLOv3, YOLOv4, and SSD300) for date palm tree detection. It is found that for the amount of training data used, the YOLOv5m (medium depth) model records the highest accuracy, resulting in a mean average precision of 92.34%. Hu *et al*. (2022) [16] present a pipeline for the monitoring and clustering of 259 peach tree crowns based on UAV images of a peach orchard in Southeast China and designed conditional generative adversarial networks (cGANs) to extract the crown area. The results of Yu *et al*. (2022) [40] showed that the Mask R-CNN model achieved the highest accuracy for individual tree detection (F_1 score = 94.68%) compared to the local maxima algorithm and marker-controlled watershed segmentation.

A great number of recent studies have proved that the integration of deep learning CNNs such as YOLOv5 and UAV images can realize the accurate detection of individual trees including fruit trees such as coconut and peach. However, to the best of our knowledge, this integration has not witnessed its application in detecting and mapping individual pomelo trees (IPTs), which are widely planted in South China (e.g., Guangdong, Guangxi, Fujian, Jiangxi, and Hunan provinces), Southeast Asia, and South America and whose fruit are popular worldwide. One reason for this is the lack of high-quality training and validation samples of IPTs since most previous studies focused on other trees or fruit trees, such as palm, olive, or coconut, and preparing these samples is time-consuming and labor-extensive. The second and also key reason is that a specific fruit tree detection model base on deep learning requires training, validation, and testing on a special dataset since deep learning is a data-driven science.

Inspired by the great progress in individual tree detection using deep learning and UAV remote sensing, we proposed an approach to establishing an accurate IFT detection model integrating deep learning with UAV remote sensing images of fruit trees to address the gap mentioned above. With this model, the location and spatial distribution of IFTs can be rapidly and accurately mapped, and the number of IFTs can also be quickly counted. We hypothesize that state-of-the-art deep learning-based methods can detect IFTs in high-resolution true-color images with low costs, high accuracy, and high efficiency. Pomelo trees are selected as the case study fruit trees and YOLOv5 is empirically determined as the deep transfer learning model for training and validation to test the hypothesis. It is desired to provide reliable and timely basic data and technical support for smart orchard management and precision agriculture development.

More specifically, three main contributions were reported in this paper. First, a new dataset of individual pomelo tree image samples (IPTIS) was created using high-resolution imagery captured by UAV remote sensing. Second, four different scales of YOLOv5 models were trained and evaluated on the new dataset to compare and select an optimal one (namely, PomeloNet) for realizing accurate and fast detection of IPTs. Third, a thematic map showing the location, spatial distribution, and the number of the IFTs in the large-scale pomelo orchards of the study area. It can provide important reference information for the precision orchard.

The remainder of the paper is organized as follows: Section 2 presents materials and our proposed approach to detecting and mapping IFTs, followed by a brief description of the experimental setting. In Section 3, experimental results and discussion are presented. Finally, Section 5 concludes the paper.

2. Materials and Methods

2.1. Study area and pomelo tree

A large pomelo orchard is selected as the experimental study area. It is located in Shishan town, Meixian district in Meizhou city (Figure 1). Meizhou city has jurisdiction over Meijiang district, Meixian district, Pingyuan county, Jiaoling county, Dabu county, Fengshun county, Wuhua county, and Xingning city, with a total land area of 15,876 km^2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Meizhou city (23° 23′~ 24° 56′ N and 115° 18′ ~ 116° 56′ E) lies in the northeast of Guangdong province, South China.

Meizhou city is located in the transition zone between south-subtropical and mid-subtropical climate zones with a typical subtropical monsoon climate. Its annual average temperature is 21.3℃, mean annual maximum temperature is 26.3℃, and mean annual minimum temperature is 17.5℃. Its annual average rainfall reaches 1,528.6 mm. The climate of the city is affected by the specific terrain of mountainous areas accounting for about 80% of the city's total land area. It presents a kind of typical characteristics of low latitude climate such as long summer, short winter, high temperature, sufficient light, and abundant and concentrated rain with the unique features of a mountainous climate such as the wide gap between cold and heat, blocked airflow, prone to drought and flood disasters, small terrain, and prominent climate. It is very suitable for various kinds of fruit trees such as pomelos to grow in mountainous Meizhou.

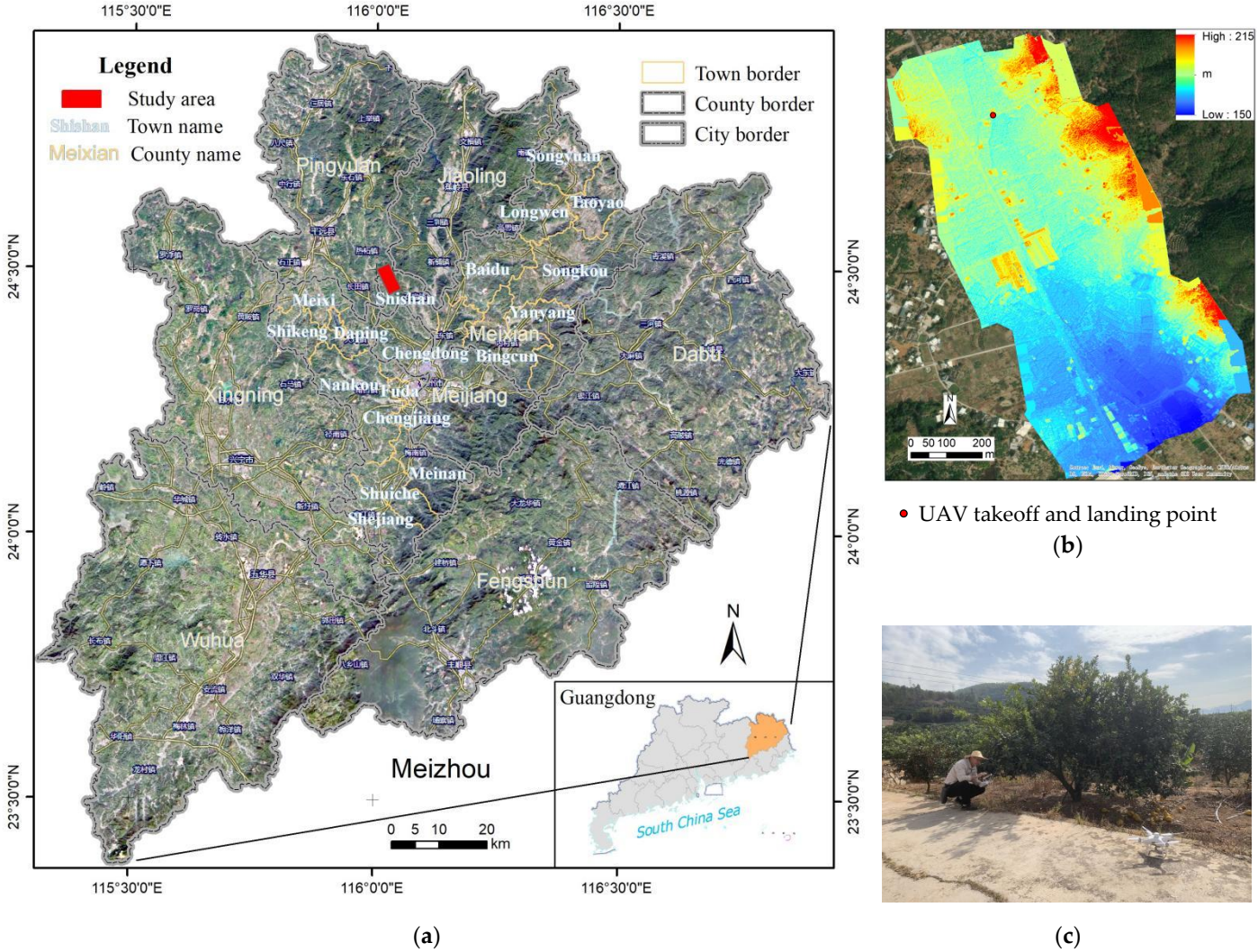


Figure 1. Overview of the experimental study area. (**a**) The location map showing the study area in a red rectangle, (**b**) The digital surface model (DSM) map, (**a**) A site photo showing the pomelo orchard and the DJI P4 drone used during the fieldwork investigation.

Pomelo (*Citrus maxima*) is a large Citrus fruit that has thick yellow skin and tastes like a grapefruit, but sweeter. It is a kind of delicious and popular fruit for its rich nutrition and high medicinal value. Pomelo tree is a medium and large-scale evergreen broad-leaved tree of Citrus in Rutaceae. A mature pomelo tree's height can reach 3-6 meters on average and its crown diameter go to 4-6 meters generally. Thus it can be recognized in high-resolution UAV-based images by visual interpretation with the assistance of field-work.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Cultivation and planting of pomelo trees in Meizhou city have undergone a long history after the fruit trees were introduced from Rong county, Guangxi, South China. After more than one century of careful improvement and cultivation, the golden pomelo fruit of Meizhou city has become very famous for its unique taste and flavor. Thus it has been designated as a national geographical indication product of China. Meizhou city was therefore dubbed the hometown of golden pomelo fruit and the golden pomelo fruit was then called Meizhou golden pomelo fruit in turn. Nowadays, all kinds of pomelos in Meizhou city are called Meizhou Pomelo Fruit (MPF), a specific name.

Meizhou city is the largest pomelo fruit-producing area in Guangdong province. The cultivation area reached 333.34 km^2 in 2018, with a total yield of 8.00×10^8 kg. Its yield accounted for 90% of the total pomelo output of Guangdong province, 20% of China, and 10% of the world. The pomelo fruit is one of the three major agricultural products in Meizhou. At present, it has become the main source of income for farmers in Meizhou. Meixian district is the largest planting area of pomelo trees among the eight sub units and is also an important golden pomelo tree planting base in Meizhou city. The planting area of pomelo trees in the Meixian district reached 160.93 km^2 in 2018, accounting for nearly half of the total planting area of pomelo trees in Meizhou, with a total yield of 4.91×10^8 kg. Shishan town is abundant in pomelo trees and convenient to access for field investigation and UAV remote sensing monitoring. The distribution of pomelo orchards in the town is relatively centralized. For this reason, it is selected as our experimental study area. However, most of the pomelo trees grow in hilly areas mixed with other trees and their crown forms and shapes are also like the surrounding trees, causing that pomelo trees can not be distinguished from some other trees in UAV images.

In September 2019, Meizhou municipal government released a plan for the MPF industry development (2019-2025) which has been the action guide for the development of the MPF industry in recent and coming years. The plan proposed to lay out the MPF industry comprehensively for the two districts, five counties, and one city in Meizhou city, striving to create an overall situation with pomelo fruit in the middle, navel orange in the north, honey pomelo fruit in the southeast, and golden pomelo fruit in the southwest, and carry on ten key projects for the development of the MPF industry in Meizhou. It aimed to make Meizhou be a nationally famous demonstration city of pomelo fruit industrialization and make MPF be one of the most famous pomelo fruit brands in China. Therefore, it is of significance to detect IPTs, map their location and spatial distribution, and count their planting area and number. It is desired to provide reference information for cultivation area investigation, yield prediction, and smart orchard management and plan in Meizhou city.

2.2. Our proposed approach for individual fruit tree detection and mapping

This study proposed an approach for detecting IFTs, mapping their spatial distribution, and counting their planting area and number by integrating deep transfer learning of YOLOv5 with high-resolution low-altitude UAV remote sensing images. The workflow of the proposed approach is illustrated in Figure 2, containing six steps shown as follows.

1. Capturing and processing UAV remote sensing images;
2. Creating a dataset of Individual Fruit Tree Image Samples (IFTIS);
3. Training, validating, and testing the selected YOLOv5 models;
4. Evaluating the accuracy and performance, an optimal YOLOv5-based model, FruitNet, for the detection of IFTs will be obtained;
5. Mapping the location and spatial distribution of IFTs using the predicted results of FruitNet;
6. Counting the planting area and the number of IFTs.

To test and validate our proposed approach, the MPF trees were selected as the example targets of fruit trees to carry on the study on individual MPF tree detection and their spatial distribution mapping. The main methods and key steps of the workflow are explained in detail in the following.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

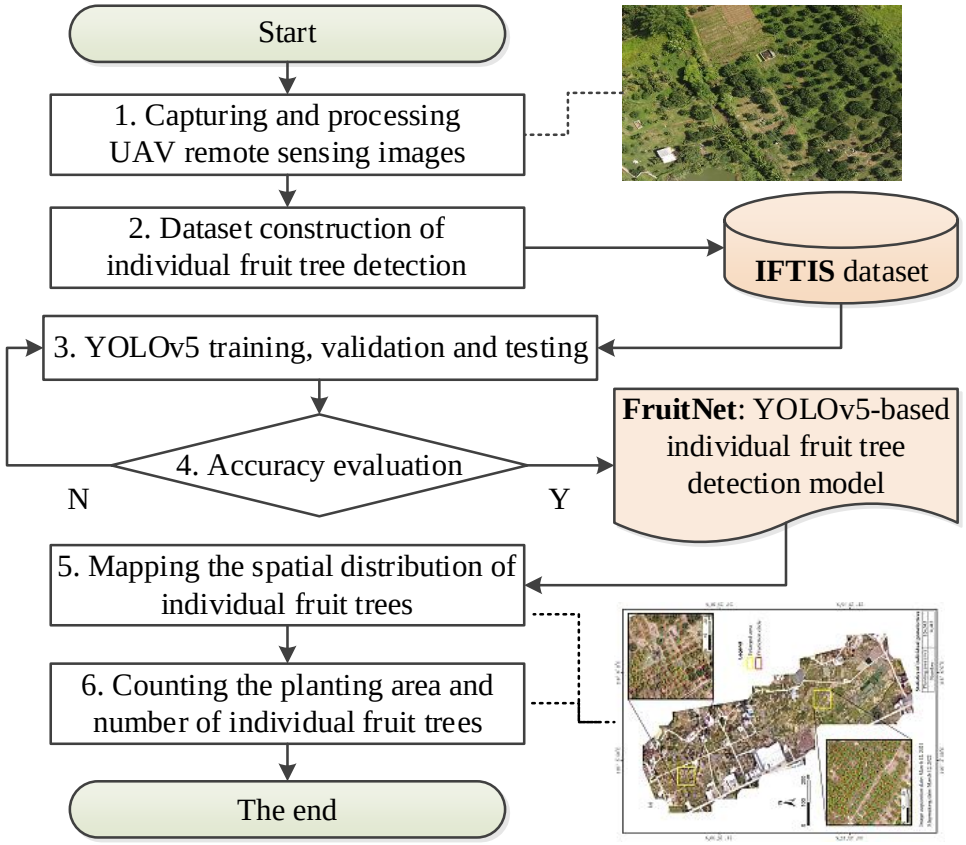


Figure 2. Workflow chart of our proposed approach to detecting and mapping individual fruit trees integrated YOLOv5 with UAV remote sensing. The dashed lines connect an original UAV image and a thematic map (rotated 90° to the left) of the present study. IFTIS denotes Individual Fruit Tree Image Samples.

2.2.1. Capturing and processing UAV remote sensing images

The DJI P4 multispectral drone was used as a UAV system to capture low-altitude remote sensing images, equipped with six 1/2.9" CMOS, including one RGB sensor for visible light imaging and five monochrome sensors for multispectral imaging (Blue, Green, Red, Red-Edge and Near-Infrared bands). It integrates RTK-enabled GNSS including GPS, BeiDou, and Galileo. So, it can capture high-quality multi-band remote sensing images without ground control points required in the traditional aerial survey. Furthermore, it can provide efficient tools for farmers in precision agriculture, greatly improving the efficiency of environmental data acquisition. To obtain high-quality UAV raw data, aerial photography tasks need to be planned before takeoff.

Pomelo trees have different morphology and spectral characteristics in different growth periods. It is particularly necessary to construct a dataset of pomelo samples consisting of UAV remote sensing images from different growth periods. In the experimental study, the same plan for three aerial photography tasks was set to capture images of the study area for three dates (i.e., February 12, March 12, and April 12, 2021). A flight altitude of 100 m was set to capture high-quality UAV raw data with a spatial resolution of 5 cm, with 60% heading and lateral overlaps. Figure 3 shows some examples of the raw true-color images collected by UAV aerial photography, which were used later to construct a dataset of RGB true-color image samples of IPTs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Figure 3. Examples of the raw true-color images captured by UAV remote sensing.

The original images obtained by UAV remote sensing on each date were preprocessed to generate a digital orthographic mosaic image model of the study area. The preprocessing steps mainly include:

i) confirming the integrity of original image data, including camera parameters in the segment and segment attributes and GNSS information;

ii) establishing engineering files and importing original image data, creating engineering, adding image data, setting image attributes, and camera model parameters in the Pix 4D Mapper software;

iii) automatic processing of the UAV images, including initialization, point cloud encryption, regional 3D reconstruction, and digital orthographic image model generation.

(**a**) (**b**) (**c**)

Figure 4. Digital mosaic orthographic images for three dates (a) February 12, 2021; (b) March 12, 2021; and (c) April 12, 2021 in the study area.

The three digital orthographic mosaic image models of the study area generated through the above processes are shown in Figure 4, showing different characteristics of color tones. It should be noted that the actual extents of the three image models have some differences although the same flight plan was set. It spent about 40 minutes, two flights of UAV aerial photography, completing the task of capturing remote sensing images in the whole study area. During this task, the battery onboard needed to be replaced one time, causing the second flight would start from a different place and thus some images of different extent areas would be captured.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

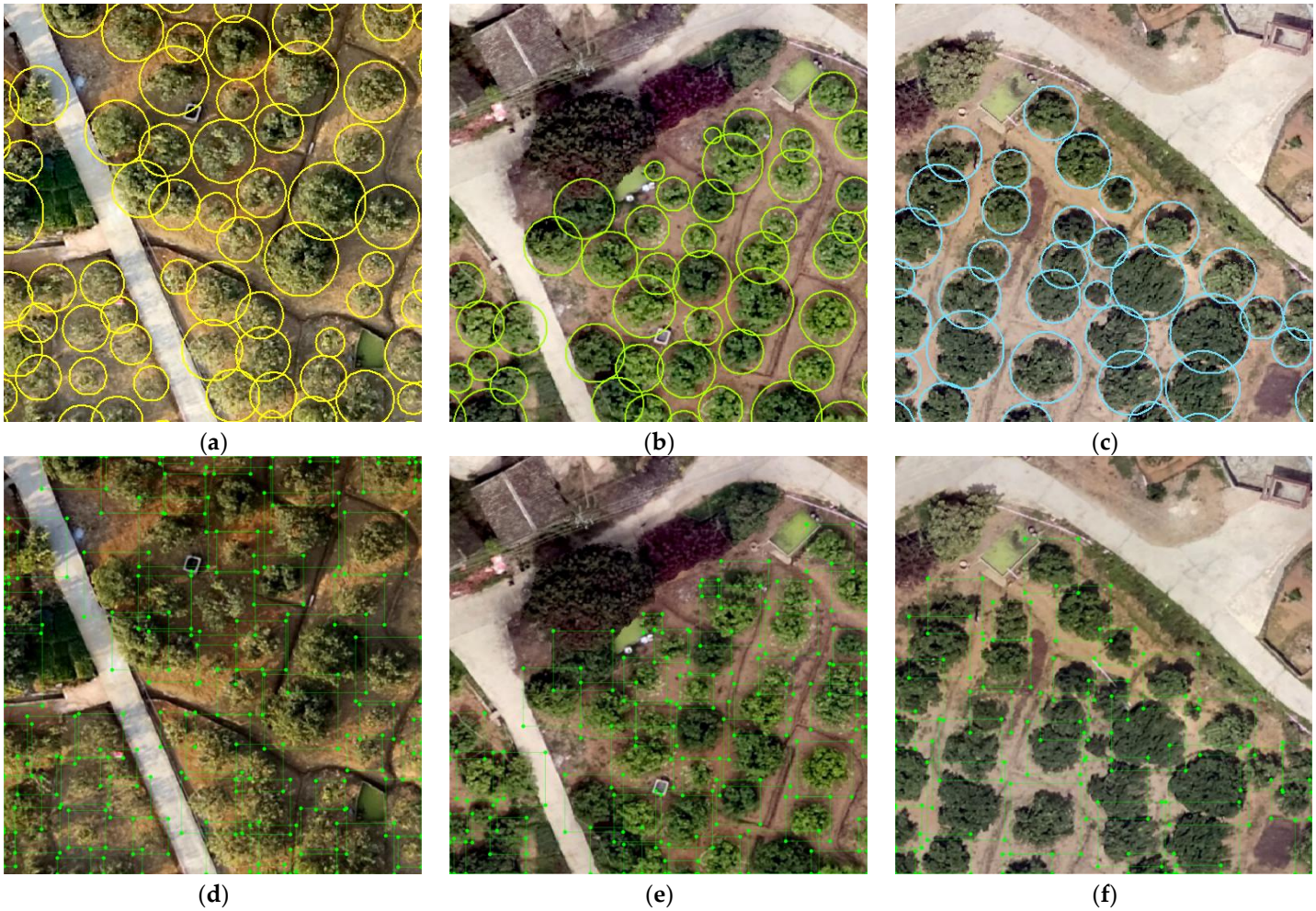


Figure 5. Annotation samples (a) February 12, 2021; (b) March 12, 2021; and (c) April 12, 2021 in ArcMap and label samples (d) February 12, 2021; (e) March 12, 2021; and (f) April 12, 2021 of the dataset of individual pomelo tree image samples in the study area.

2.2.2. Construction of a dataset of individual pomelo tree image samples

The processed mosaic orthographic images of pomelo trees were imported into ArcMap 10.8, and the ArcMap deep learning module was used to label individual pomelo tree samples. After labeling, cropping, and exporting, an individual pomelo tree detection dataset based on UAV remote sensing images was generated and named the dataset of Individual Pomelo Tree Image Samples (IPTIS). The steps are as follows. Firstly, we created three shapefiles of the surface element class vectors for different months in ArcMap, drew circle elements for the pomelo sample annotations manually according to the records of field investigation, added class files in the properties table of surface element class vector, and identified the individual pomelo tree sample's category (i.e., 1). Three annotation examples are shown in Figures 5a, 5b, and 5c. Secondly, the polygon feature-class files were used to export the images and their corresponding annotated sample data, suitable for the subsequent research requirements. The digital orthographic image of the study area taken each month was cropped into clip images with the size of 640 × 640 and zero overlaps. The images without pomelo tree annotations were excluded when exporting in ArcMap. Lastly, a dataset of IPTIS was created according to the PASCAL VOC [41] data format by combining all exported clip images for three months, with a total of 439 images. Three label examples of the clip images of the dataset obtained after cropping and exporting are shown in Figures 5d, 5e, and 5f. The actual label of an individual pomelo tree is the minimum bounding rectangle of the drawn circle which will be the ground truth for model training and validation in deep learning.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

2.2.3. YOLOv5 deep learning object detection model

Based on the dataset IPTIS, the single-stage object detection algorithms of YOLOv5 (You Only Look Once) were empirically selected to train, validate, and test models for individual pomelo tree detection. YOLOv5 is the fifth version of the YOLO model family and has been widely used in object detection tasks such as pedestrians, vehicles, and ships. The YOLO model is divided into three parts: backbone network (Backbone), neck network (Neck), and head network (Head) (Figure 6). A backbone network is used to extract features from the input data; the neck network collects and distributes features of different scales; the head network is used to judge the positioning and category of the target box. In YOLOv5, the backbone network adopts a cross-stage local network [42] to solve the problem of gradient information duplication and gradient disappearance of network optimization; it adopts a path aggregation network [43] and spatial pyramid pooling network [44]. As a neck network, the model enhances the detection of objects with different scaling scales to identify the same object of different scales; the head network uses the same detection layer as YOLOv3 and YOLOv4, applies the best anchor box to the feature map, and generates the final output vector with category probability, object score, and prediction bounding box.

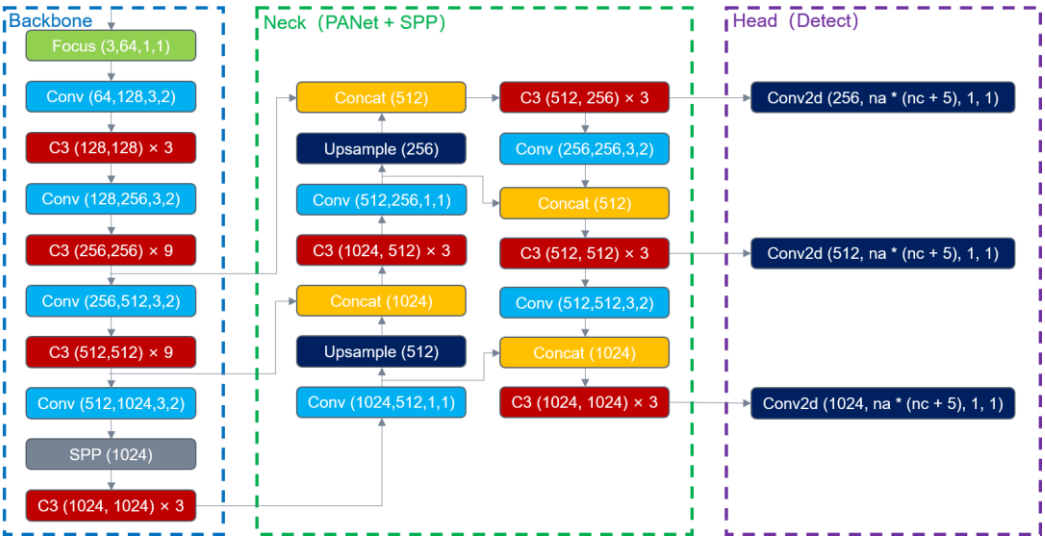


Figure 6. The YOLOv5 model structure.

Taking advantage of the configuration flexibility of the YOLOv5 model, the depth_multiple and width_multiple parameters in the model structure file were modified specifically. The model scale was indirectly controlled by adjusting the number of bottlenecks in CSPs and the number of convolution cores in each convolution layer, respectively. Four YOLOv5 models with different scales were obtained (Table 1), namely YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLO5x (extreme). Their layers, parameters, and GFLOPS (Giga Floating-point Operations Per Second) increase with the rise of depth and width, and the complexity of the model increases accordingly.

Three steps were implemented to obtain an optimal detection model based on YOLOv5. First, the four different scales of YOLOv5 models were used for training and validation. The model with the highest accuracy was selected as IPT detection model according to the comparison results of the evaluation metrics. This preferred model was named PomeloNet for short. Second, the detection ability of the individual pomelo tree detection model in other areas was tested to compare the robustness and generalization abilities of the model. Finally, PomeloNet was used to detect IPTs in the experimental area. and its spatial distribution map was made.

Table 1. Parameters of various scales of YOLOv5 models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Model	Depth	Width	Number of layers	Number of parameters	GFLOPS [1]
YOLOv5s	0.33	0.50	283	7.3 M	17.1
YOLOv5m	0.67	0.75	391	21.4 M	51.4
YOLOv5l	1.00	1.00	499	47.1 M	115.6
YOLOv5x	1.33	1.25	607	87.8 M	219.0

[1] GFLOPS denotes Giga Floating-point Operations Per Second.

2.2.4. Evaluation metrics

In deep learning, the evaluation of models is very important. Only by selecting the appropriate evaluation method can we quickly find out the possible problems of the model in the training process and find a suitable method to optimize the model. The confusion matrix is not only a standard format for evaluating accuracy but also a visualization tool capable of using special matrices to present the effect of model performance. The confusion matrix only consists of positive and negative examples. Table 2 shows the confusion matrix for a classic example of binary classifications, where each column represents the predicted value and each row represents the actual category.

Table 2. Confusion matrix of binary classification of artificial intelligence.

Confusion matrix		Predicted label	
		true	false
actual label	positive	TP *	FP
	negative	TN	FN

* TP means True Positive, signifying that the actual category of the sample is positive, and the result predicted by the model is also positive. TN means True Negative, signifying that the actual category of the sample is negative, and the model predicts it to be negative. FP means False Positive, signifying that the actual category of the sample is negative, but the model predicts it to be positive. FN means False Negative, signifying that the actual category of the sample is positive, but the model predicts it as negative.

The present study is an example of binary classifications. The precision, recall, F_1 score, and average precision metrics were used to evaluate the accuracy of the trained YOLOv5 models.

(1) Precision and recall

According to Table 2, precision (P) and recall (R) metrics are defined as Equations 1 and 2, respectively. The precision denotes the proportion of actual positive samples among all the results predicted as positive samples. The recall denotes the proportion of the samples predicted as positive examples by the classifier to the actual number of positive examples, also known as sensitivity, describing the classifier's sensitivity to the category of positive examples.

$$P = TP / (TP + FP), \quad (1)$$

$$R = TP / (TP + FN), \quad (2)$$

where, P and R denote the precision and recall, respectively; TP, FP, and FN indicate the same meanings as in Table 2.

(2) F_1-score

The F_1-score is the harmonic mean of precision and recall, taking both metrics into account in Equation 4.

$$F_1 = 2 \times P \times R / (P + R), \quad (3)$$

where, F_1 denotes the F_1-score; P and R denote the precision and recall, respectively.

(3) Average precision

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

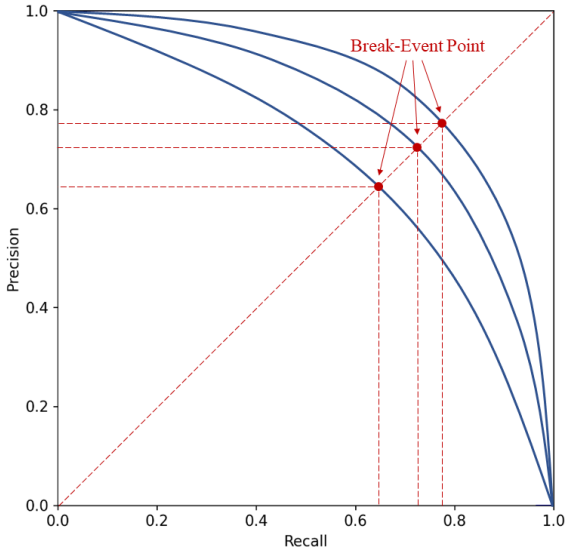


Figure 7. Precision-recall curves.

In the field of deep learning object detection, the Average Precision (AP) measures how well a model detects a specific category, represented by a Precision-Recall Curve (PRC) graph (Figure 7). The PRC graph is a chart with horizontal recall and vertical precision, a monotonous decreasing curve. The area below the PR curve of a specific category is defined as the AP, as defined in Equation 5:

$$AP = \int_0^1 f(c)d(c) \tag{4}$$

where, AP represents the average precision, $f(c)$ means the precision-recall curve of category c. In the PR plot, the closer the curve is to the upper right corner, the higher the model's accuracy. In addition to using the area evaluation model below the curve, we can also draw a line with a slope of 1 on the PRC graph, and the intersection point of the line and the PR curve is the equilibrium point F_1. The score is an F_1 score.

The mean Average Precision (mAP) used to obtain the average accuracy of multiple categories, is a comprehensive measure of a model for all detection categories. There is only one pomelo tree category excluding the background in this study, so mAP = AP.

(4) Intersection over union

In object detection, the ability of the representation model is not only the prediction probability of category but also the positioning accuracy of the prediction box. The ratio of Intersection over Union (IoU) is commonly used as a matching degree evaluation metric of the prediction bounding box and the ground truth box in the data set (Figure 8), calculated by their area intersection and union ratio (Equation 6). The higher the ratio value, the better the matching degree. For an ideal result, the prediction box and ground truth box overlap completely, and the ratio value reaches 1.

$$IoU = \frac{Area(B) \cap Area(G)}{Area(B) \cup Area(G)} \tag{5}$$

where, $Area(B)$ represents the area of the prediction bounding box, and $Area(G)$ represents the area of the ground truth box.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

Figure 8. Intersection over the union of ground truth and prediction bounding box.

The threshold standard for positive cases is IoU > 0.5, otherwise negative cases. So, the AP@0.5 used below denotes the average precision when the IoU > 0.5, and the AP@0.5:0.95 used below represents the average precision when the IoU lies between 0.5 and 0.95. In addition, the inference time is also an important indicator to evaluate the model's ability in object detection. Frames Per Second (FPS) are usually used to measure the model inference speed.

2.2.5. Mapping spatial distribution and counting the planting area and number of IPTs

Due to the limitations of the device graphic card's memory and the image size of the input data of the model, the high-resolution UAV remote sensing image of the whole area cannot be directly predicted and detected. Etten (2018) [12] proposed a YOLT method based on YOLOv2 to apply the object detection model to large-scale remote sensing images. Therefore, according to the post-processing idea of YOLT, the whole digital orthographic image of the study area was cropped with slicing based on a specific image size of a model and an overlap degree of 20%. The cropped images were named the following:

ImageName_row_column_width_height_0_globe^row_globe^column.jpg

where, ImageName represents the name of the original image; row and column represent the coordinates of the upper left corner of the slice in the original image; width and height represent the slice size, and globe^row and globe^column represent the number of rows and columns of the original image matrix, respectively, namely the original image size. For example, for the 202102_0_0_640_640_0_11227_16272.jpg, the original image name is 202102. The slice's upper left corner corresponds to the original image coordinates (0, 0), the slice size is 640 × 640, and the original image size is 11,227 × 16,272. This can help post-processing to tessellate the slice images in the following steps.

Four steps were adopted to make the spatial distribution map of IPTs. Firstly, the YOLOv5x model was used to detect IPTs and obtain the coordinates of the detected trees in the slice images. Then, the repeatedly detected target boxes in the slice overlap area were deleted by using the non-maximum suppression method based on the overlap degree of 20% of the neighboring slices. Thirdly, the slice images were tessellated into a large-scale image according to the slice name, and the remaining target boxes were overlaid on the tessellated image. Finally, a thematic map of the detection results of the IPTs and their spatial distribution was made using the ArcMap software. The number and planting area of the detected IPTs in the study area were also counted in the ArcMap software, which can be added to the thematic map as statistic annotation information.

2.3. Experimental environment and setup

The experimental platform was configured with the following:

- Intel (R) Core (TM) i3-10100F CPU @ 3.60 GHz processor,
- NVIDIA GeForce RTX 2080 Ti 11 GB GPU independent graphics card,
- 64-bit Linux-5.4.0 Debian,
- Python 3.7,
- PyTorch 1.8.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

By cross-validation, 395 (90%) images were randomly selected as training set and 22 (5%) images as validation set; the remaining 22 (5%) images were used as the test set to test the final model. The deep transfer learning strategy was used to train on the pomelo image dataset to obtain the deep learning model for the detection of IPTs. The MS-COCO pretrained model weights of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x were used to retrain on the training set. The hyperparameters were finetuned to obtain the optimal ones, shown as follows.

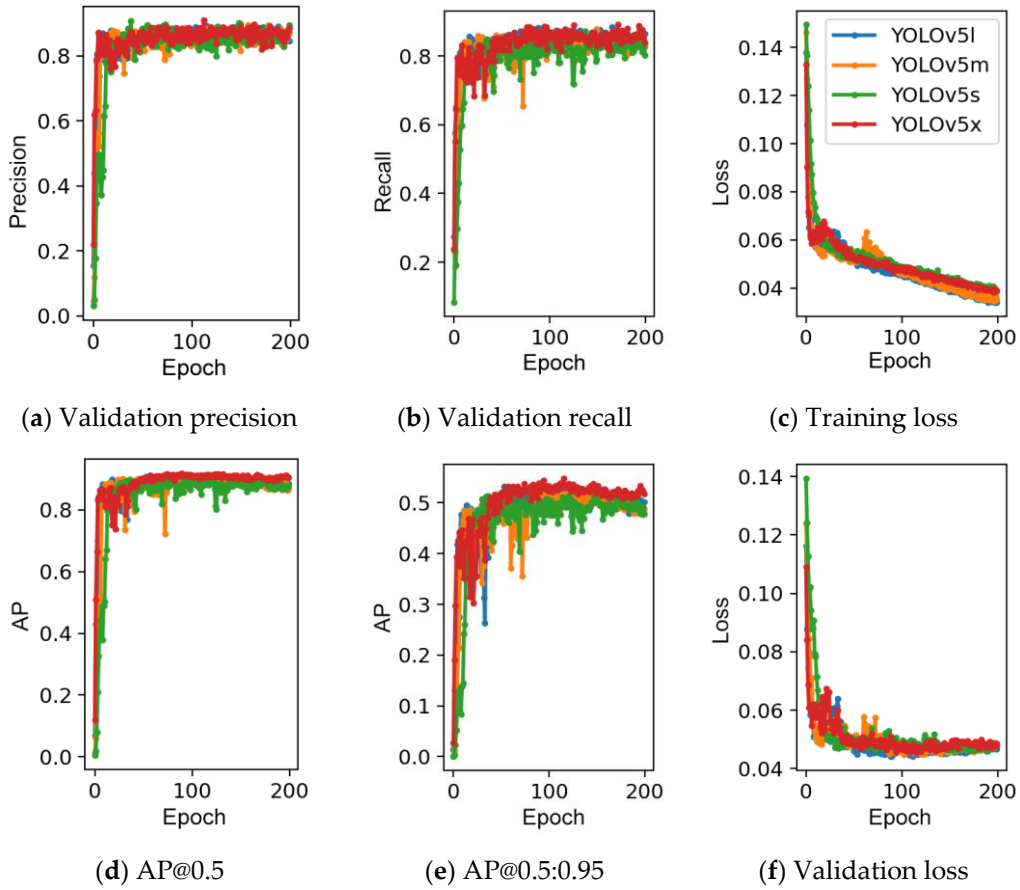
- Max epoch: 200,
- Minimum batch size: 32,
- Initial learning rate: 0.0001,
- Optimizer: SGDM.

The max epoch was set to 200 and the minimum batch size was set to 32 to keep the memory utilization of GPU at about 90% and get good accuracy. The initial learning rate was set to 0.0001 and the models were trained using a Stochastic Gradient Descent with Momentum (SGDM) strategy to avoid overfitting and underfitting.

3. Results and discussion

3.1. Comparison of accuracies of the four YOLOv5 models

As shown in Figure 9, the precision, recall, and average accuracy of the four YOLOv5 models almost stabilize after 100 epochs. To avoid the fluctuating influence of the evaluation metrics, the average value of each model's accuracy trained within 101~200 epochs was selected as the final evaluation value.



(**a**) Validation precision (**b**) Validation recall (**c**) Training loss

(**d**) AP@0.5 (**e**) AP@0.5:0.95 (**f**) Validation loss

Figure 9. Visualization of the evaluation metrics for the four YOLOv5 models.

Figure 10 shows the average precision (AP@0.5, hereinafter AP) variation of each YOLOv5 model during the training process from 20 to 100 epochs. With the help of the pre-trained model weights, each model can achieve high accuracy quickly. The AP of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

YOLOv5s, YOLOv5m, and YOLOv5l all achieve more than 0.83 after 20 epochs of training, while YOLOv5l even reaches more than 0.88. With the increase of iterative training times, the AP of each model is also gradually improving. After 73 epochs of training, the AP of YOLOv5x exceeds the other three models and then reaches the fitting state and remains stable.

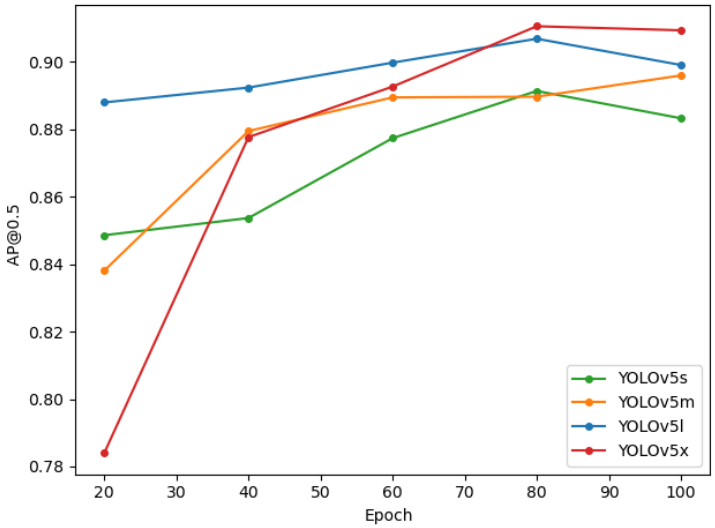


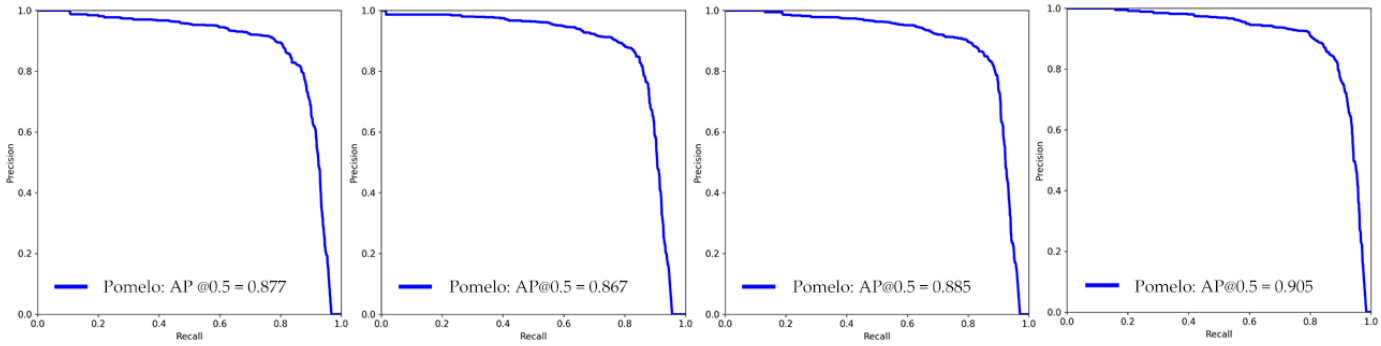
Figure 10. Average precision (AP) variation of each YOLOv5 model (20~100 epochs).

By comparing the accuracy evaluation metrics of the four models in Table 3, it can be found that as the scale of the model expands, the precision, recall, F_1 score, and AP increase slightly, achieving relatively high accuracy. All of the four accuracy metrics of the YOLOv5x model are higher than those of the other three models, suggesting that YOLOv5x is the best model of accuracy performance.

Table 3. Results of the evaluation metrics of various scales of YOLOv5 models.

Models	Precision	Recall	F_1-score	AP@0.5 [1]	AP@0.5:0.95
YOLOv5s	0.857	0.824	0.840	0.878	0.492
YOLOv5m	0.862	0.839	0.850	0.885	0.501
YOLOv5l	0.867	0.849	0.858	0.891	0.507
YOLOv5x	0.869	0.855	0.862	0.907	0.524

[1] AP@0.5 means the average precision when the IoU > 0.5 and the AP@0.5:0.95 denotes the average precision when the IoU lies between 0.5 and 0.95.



(**a**) YOLOv5s (**b**) YOLOv5m (**c**) YOLOv5l (**d**) YOLOv5x

Figure 11. Precision-recall (PR) curve graph of the four YOLOv5 models.

The comprehensive performance of each model can be represented by PR plots (Figure 11). From Equation 4, the detection accuracy of the IPTs can be expressed by the area

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

below the PR curve of the pomelo tree category. Considering only the model's accuracy without real-time detection, YOLOv5x is the best model with the highest AP of 0.905 and was selected as the model for the subsequent inference and mapping.

3.2. Training and validation losses of the four YOLOv5 models

The loss indicators during the training and validation processes from 101 to 200 epochs are shown in Table 4. If the training loss value is close to the value of the validation loss, the model is not overfitting. The lower the loss, the better the accuracy of the model. Therefore, none of the YOLOv5 models was overfitting during the training on the individual pomelo tree detection dataset. YOLOv5x has the least validation loss, implying that it has the best performance among the four models.

Table 4. Training and validation losses of various scales of YOLOv5 models.

Models	Training loss	Validation loss
YOLOv5s	0.070	0.070
YOLOv5m	0.064	0.069
YOLOv5l	0.063	0.069
YOLOv5x	0.069	0.066

3.3. Training time and inference speeds of the four YOLOv5 models

As shown in Table 5, with the scale growth of YOLOv5, the training time becomes longer and the inference speed becomes shorter. All models can achieve real-time inference (i.e., FPS are greater than 60) except YOLOv5x. YOLOv5x can predict in near-real time. The YOLOv5s model spends the shortest time on training and predicts the fastest with the same amount of image data.

Table 5. Training time and inference speed of various scales of YOLOv5 models.

Models	Training time	Inference speed (FPS [1])
YOLOv5s	16 m 51 s	109
YOLOv5m	31 m 59 s	86
YOLOv5l	49 m 09 s	72
YOLOv5x	01 h 30 m 26 s	50

[1] FPS denotes Frames Per Second.

3.4. Statistics and mapping of the individual pomelo trees

A thematic map (Figure 12) was made to show the spatial distribution of the detected IPTs using the retrained YOLOv5x model, i.e., PomeloNet. As shown in the two inserted square boxes of Figures 12b and 12c, it can be found that different sizes of IPTs are almost detected accurately, indicating that PomeloNet could have a high enough accuracy to complete the task of individual pomelo tree detection. It is necessary to further test and validate the feasibility of the application of PomeloNet in other pomelo orchards.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

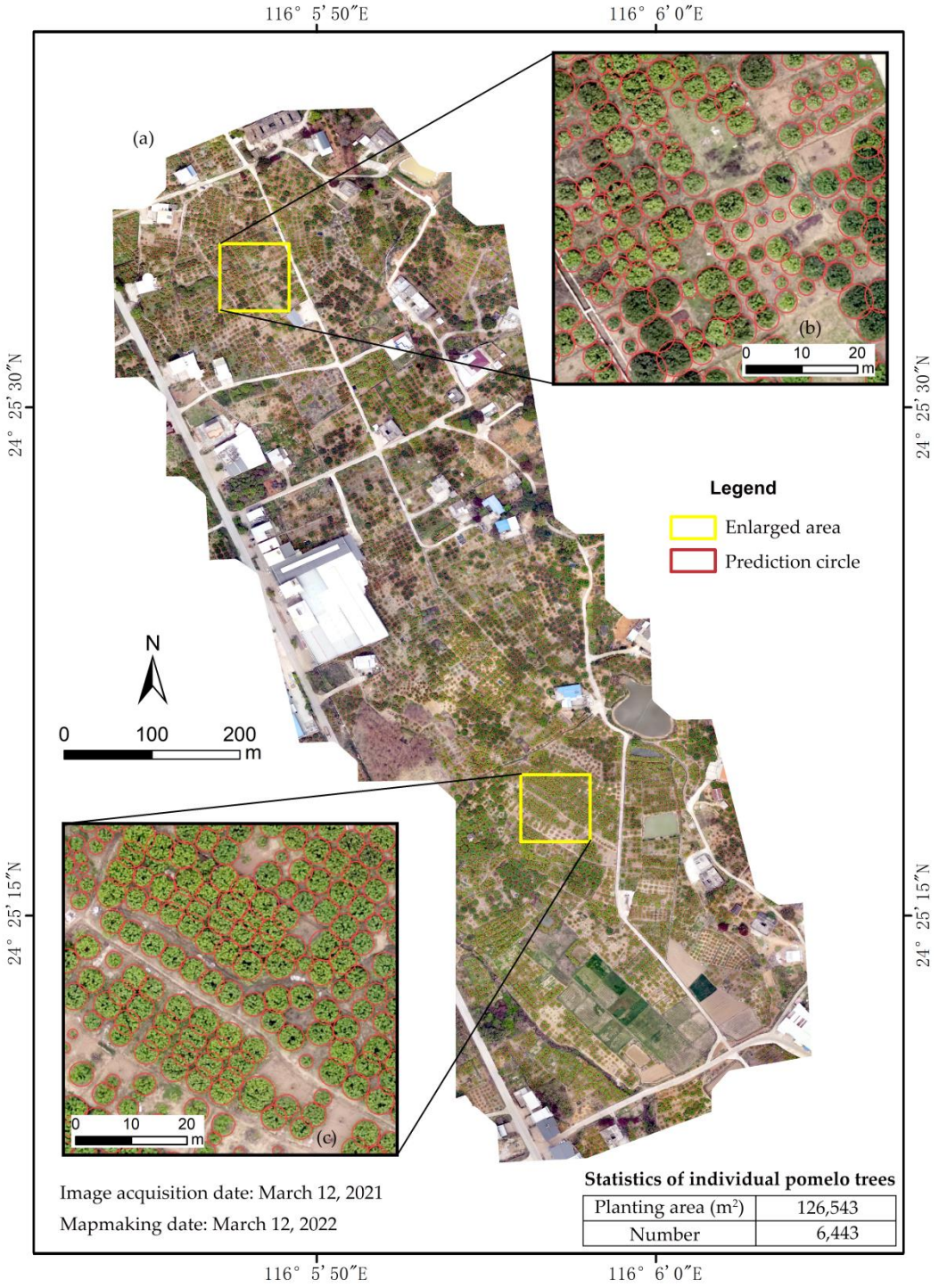


Figure 12. A thematic map showing the spatial distribution of the detected individual pomelo trees using YOLOv5x and the mosaic image acquired on March 12, 2021, and their planting area and number information with two enlarged square regions inserted.

The original square labels detected by using PomeloNet were converted into the corresponding circles to reduce the overlapping effect in the map. The planting area and the number of the detected IPTs in the experimental study area were counted with the ArcMap software and shown in the thematic map. The total planting area is summed by the area of each of IFTs. The results show that these two values are 126,543 m² and the total 6,443, respectively.

3.5. PomeloNet validation and test on images captured in the other two pomelo orchards

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

We collected UAV based images in the other pomelo orchards in Shishan town in different months to validate and test the PomeloNet's robustness and generalization performance. The four selected sample clip images were acquired in July 5, 2021 and have not been manually annotated. Hence, we cannot evaluate the accuracy and performance quantitatively through validation and test but qualitatively via visual estimation. As shown in Figure 13, most IPTs (probably > 85%) in the four images can be truly predicted although a few IFTs can not be detected or be confused by other trees. Furthermore, PomeloNet can truly detect both densely and sparsely distributed IPTs of different sizes with relatively high confidence (Figure 13). The background conditions such as other kinds of trees or shrubs may badly affect the detection results. It is necessary to append more high-quality images to the IPTIS dataset and make more quantitative accuracy and performance evaluation with more different seasons of images across various pomelo orchards to set up a reliable basis for detection of IPTs in a larger region even a county.

(**a**) (**b**) (**c**) (**d**)

Figure 13. Four sample clip images overlaid with detected boxes and confidences predicted by PomeloNet. (a) and (b) are UAV based RGB images captured by DJI P4 multispectral sensors with a flight height of 100 m; (c) and (d) are UAV based RGB images captured by DJI Phantom 4 RTK with a flight height of 120 m. The four images were all captured on July 16, 2020, but look different in color for capturing via different sensors and in different orchards.

3.6 Limitations and future work

Despite a lot of hard work, there are some limitations in dataset creation, deep learning model selection and design, and hyperparameter optimization. First, although we acquired both RGB and multispectral images using UAV remote sensing, only the UAV-based RGB images were used to construct the IPTIS dataset in the present study. The UAV-based multispectral images will be used in future study with the hope of improving the accuracy of the models. Other UAV-based high-resolution images such as hyperspectral or LiDAR imagery would be better options for the detection of IPTs because their more spectral information or highly effective point cloud data [16, 17] could reveal more detailed features and improve the performance of CNNs that helps distinguish IPTs from the images. Second, more CNN models such as Faster R-CNN, U-Net, SDD, and Mask R-CNN [1, 38-40] should be trained and tested for obtaining a better model to fulfill the task. The structures of the selected models could even be modified to improve accuracy and performance [14] for better precise applications of smart orchard management. Third, data augmentation and hyperparameter optimization need to be further carried out for obtaining a more robust model of performance. These all deserve further research.

In the future, these kinds of spatial and attribute data about the individual fruit trees (i.e., distribution, planting area, and number) in an orchard (for example, a pomelo orchard) can be obtained through our proposed approach. These data could be easily integrated into a smart orchard management system that could provide fast growth monitoring of individual fruit trees, accurate yield estimation of the fruit, real-time disease prevention and control, and precision cultivation and management. Town-level even county- and cit thematic maps of IPTs will be made through our proposed approach in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

the coming study. The pomelo yield estimation based on the thematic map of IPTs will be an important topic in our future research.

4. Conclusions

In the present study, we proposed a deep learning approach to detecting and mapping individual fruit trees in UAV remote sensing imagery, taking the pomelo trees in Meizhou city as an experimental study example. UAV remote sensing technology was applied to acquire high spatial-resolution images of the study area. These images were preprocessed in the Pix 4D Mapper software. A dataset of individual pomelo tree image samples (IPTIS) was constructed through visual interpretation and the deep learning tools in the ArcGIS software combined with fieldwork investigation. Four different scales of YOLOv5 (i.e., YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) object detection models were used to train and validate on the dataset. The evaluation results show that all four models achieve relatively high detection accuracy. Among the four models, YOLOv5x performed outstandingly with the highest accuracy. The retrained YOLOv5x model, namely PomeloNet, was thus selected to detect and post-process IPTs in the whole mosaic orthographic image of the study area. Finally, a spatial distribution thematic map of IPTs was made according to the detection results. This study provides reference information for related research and smart orchard management.

Author Contributions: Conceptualization, Yongzhu Xiong and Xiaofeng Zeng; Data curation, Xiaofeng Zeng and Yankui Chen; Formal analysis, Yongzhu Xiong, Xiaofeng Zeng and Mingyong Zhu; Funding acquisition, Yongzhu Xiong, Xiaofeng Zeng and Mingyong Zhu; Investigation, Yongzhu Xiong, Yankui Chen and Mingyong Zhu; Methodology, Yongzhu Xiong and Xiaofeng Zeng; Project administration, Yongzhu Xiong and Mingyong Zhu; Resources, Yankui Chen; Software, Yongzhu Xiong, Xiaofeng Zeng and Weiqian Lai; Supervision, Yongzhu Xiong; Validation, Xiaofeng Zeng, Jiawen Liao and Weiqian Lai; Visualization, Jiawen Liao and Weiqian Lai; Writing – original draft, Yongzhu Xiong and Xiaofeng Zeng; Writing – review & editing, Yongzhu Xiong, Xiaofeng Zeng and Mingyong Zhu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangdong Province Special Project in Key Fields for Universities (New Generation Information Technology), grant number 2020ZDZX3044, and the Guangdong Provincial Science and Technology Innovation Strategy Special Fund ("Climbing Plan" Special Fund) Project, grant number PDJH2020b0552. This study was partly supported by the Ordinary University Characteristic Innovation Project of Guangdong Province, grant number 2020KTSCX140, the Research Ability Improvement Project of Key Construction Disciplines in Guangdong Province, grant number 2021ZDJS073, and the Intangible Cultural Heritage Research Base Project of Guangdong Province, grant number 17KYKT13.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The IPTIS dataset used during the current study is available from the corresponding author (Y.X.) upon reasonable request.

Acknowledgments: Y.X. would like to acknowledge the support from the China Scholarship Council (Grant number 201808440171), Guangdong Provincial Key Laboratory of Conservation and Precision Utilization of Characteristic Agricultural Resources in Mountainous Areas (Grant number 2020B121201013) and Guangdong Pomelo Engineering Technology Development Center (Grant number 2019GCZX007). The authors would like to thank Prof. Liangzheng Xu, Prof. Guangrui Zhong, Prof. Kekun Huang, Dr. Zhuolun Xie, and Mr. Changbi Zhu at the Jiaying University, for their great help in pomelo orchard selection and image acquisition and discussions on experimental design. The authors would like to thank the editors and reviewers for their valuable comments and suggestions that greatly improve our manuscript and the *Research Square* for posting our manuscript preprint online at https://doi.org/xxxxxx.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

References

1. Jintasuttisak, T.; Edirisinghe, E.; Elbattay, A., Deep neural network based date palm tree detection in drone imagery. *Comput Electron Agr* **2022**, 192, 106560.
2. Li, D.; Li, M., Research Advance and Application Prospect of Unmanned Aerial Vehicle Remote Sensing System. *Geomat Infor Sci Wuhan Univ* **2014**, 39, (5), 505-513.
3. Colomina, I.; Molina, P., Unmanned aerial systems for photogrammetry and remote sensing: A review. *Isprs J Photogramm* **2014**, 92, 79-97.
4. Khanal, S.; Fulton, J.; Shearer, S., An overview of current and potential applications of thermal remote sensing in precision agriculture. *Comput Electron Agr* **2017**, 139, 22-32.
5. Ghali, R.; Akhloufi, M.A.; Mseddi, W.S., Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation. *Sensors-Basel* **2022**, 22, (5), 1977.
6. Immerzeel, W.W.; Kraaijenbrink, P.D.A.; Shea, J.M.; Shrestha, A.B.; Pellicciotti, F.; Bierkens, M.F.P.; de Jong, S.M., High-resolution monitoring of Himalayan glacier dynamics using unmanned aerial vehicles. *Remote Sens Environ* **2014**, 150, 93-103.
7. Wu, Y.; Shan, Y.; Lai, Y.; Zhou, S., Method of calculating land surface temperatures based on the low-altitude UAV thermal infrared remote sensing data and the near-ground meteorological data. *Sustain Cities Soc* **2022**, 78, 103615.
8. Osco, L.P.; Marcato Junior, J.; Marques Ramos, A.P.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J., A review on deep learning in UAV remote sensing. *Int J Appl Earth Obs* **2021**, 102, 102456.
9. Zhang, L.; Zhang, L.; Du, B., Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *Ieee Geosc Rem Sen M* **2016**, 4, (2), 22-40.
10. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J., In *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*, Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23-28 June 2014, 2014; Columbus, OH, USA, 2014; pp. 580-587.
11. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A., In *You Only Look Once: Unified, Real-Time Object Detection*, Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016, 2016; Las Vegas, NV, USA, **2016**; pp. 779-788.
12. Etten, A.V., You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery. *arXiv:1805.09512* **2018**.
13. Yan, J.; Zhao, Y.; Zhang, L.; Su, X.; Liu, H.; Zhang, F.; Fan, W.; He, L., Recognition of Rosa roxbunghii in natural environment based on improved Faster RCNN. *Trans Chin Soc Agr Eng* **2019**, 35, (18), 143-150.
14. Liu, J.; Wang, X., Tomato disease and pest detection algorithm based on YOLO convolutional neural network. *China Cuc Veg* **2020**, 33, (9), 18-22+38.
15. Sun, Y.; Li, Z.; He, H.; Guo, L.; Zhang, X.; Xin, Q., Counting trees in a subtropical mega city using the instance segmentation method. *Int J Appl Earth Obs* **2022**, 106, 102662.
16. Hu, J.; Zhang, Y.; Zhao, D.; Yang, G.; Chen, F.; Zhou, C.; Chen, W., A robust deep learning approach for the quantitative characterization and clustering of peach tree crowns based on UAV images. *Ieee T Geosci Remote* **2022**, 1-14.
17. Jaskierniak, D.; Lucieer, A.; Kuczera, G.; Turner, D.; Lane, P.N.J.; Benyon, R.G.; Haydon, S., Individual tree detection and crown delineation from Unmanned Aircraft System (UAS) LiDAR in structurally complex mixed species eucalypt forests. *Isprs J Photogramm* **2021**, 171, 171-187.
18. Liu, H.; Dong, P.; Wu, C.; Wang, P.; Fang, M., Individual tree identification using a new cluster-based approach with discrete-return airborne LiDAR data. *Remote Sens Environ* **2021**, 258, 112382.
19. Xu, X.; Zhou, Z.; Tang, Y.; Qu, Y., Individual tree crown detection from high spatial resolution imagery using a revised local maximum filtering. *Remote Sens Environ* **2021**, 258, 112397.
20. Yun, T.; Jiang, K.; Li, G.; Eichhorn, M.P.; Fan, J.; Liu, F.; Chen, B.; An, F.; Cao, L., Individual tree crown segmentation from airborne LiDAR data using a novel Gaussian filter and energy function minimization-based approach. *Remote Sens Environ* **2021**, 256, 112307.
21. Mäyrä, J.; Keski-Saari, S.; Kivinen, S.; Tanhuanpää, T.; Hurskainen, P.; Kullberg, P.; Poikolainen, L.; Viinikka, A.; Tuominen, S.; Kumpula, T.; Vihervaara, P., Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens Environ* **2021**, 256, 112322.
22. Yin, T.; Zeng, J.; Zhang, X.; Zhou, X., Individual Tree Parameters Estimation for Chinese Fir (Cunninghamia lanceolate (Lamb.) Hook) Plantations of South China Using UAV Oblique Photography: Possibilities and Challenges. *Ieee J-Stars* **2021**, 14, 827-842.
23. Luo, H.; Khoshelham, K.; Chen, C.; He, H., Individual tree extraction from urban mobile laser scanning point clouds using deep pointwise direction embedding. *Isprs J Photogramm* **2021**, 175, 326-339.
24. Yao, L.; Liu, T.; Qin, J.; Lu, N.; Zhou, C., Tree counting with high spatial-resolution satellite imagery based on deep neural networks. *Ecol Indic* **2021**, 125, 107591.
25. 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 Agr* **2020**, 174, 105504.
26. Weinstein, B.G.; Marconi, S.; Bohlman, S.A.; Zare, A.; White, E.P., Cross-site learning in deep learning RGB tree crown detection. *Ecol Inform* **2020**, 56, 101061.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0007.v2

27. Pleșoianu, A.; Stupariu, M.; Șandric, I.; Pătru-Stupariu, I.; Drăguț, L., Individual Tree-Crown Detection and Species Classification in Very High-Resolution Remote Sensing Imagery Using a Deep Learning Ensemble Model. *Remote Sens-Basel* **2020**, 12, (15), 2426.
28. Culman, M.; Delalieux, S.; Van Tricht, K., Individual Palm Tree Detection Using Deep Learning on RGB Imagery to Support Tree Inventory. *Remote Sens-Basel* **2020**, 12, (21), 3476.
29. Zheng, J.; Fu, H.; Li, W.; Wu, W.; Zhao, Y.; Dong, R.; Yu, L., Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network. *Isprs J Photogramm* **2020**, 167, 154-177.
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. *Forest Ecol Manag* **2020**, 475, 118397.
31. Brandt, M.; Tucker, C.J.; Kariryaa, A.; Rasmussen, K.; Abel, C.; Small, J.; Chave, J.; Rasmussen, L.V.; Hiernaux, P.; Diouf, A.A.; Kergoat, L.; Mertz, O.; Igel, C.; Gieseke, F.; Schöning, J.; Li, S.; Melocik, K.; Meyer, J.; Sinno, S.; Romero, E.; Glennie, E.; Montagu, A.; Dendoncker, M.; Fensholt, R., An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **2020**, 587, (7832), 78-82.
32. Hanan, N.P.; Anchang, J.Y., Satellites could soon map every tree on Earth. *Nature* **2020**, 587, 42-43.
33. Weinstein, G.B.; Marconi, S.; Bohlman, S.; Zare, A.; White, E., Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sens-Basel* **2019**, 11, (11), 1309.
34. Freudenberg, M.; Nölke, N.; Agostini, A.; Urban, K.; Wörgötter, F.; Kleinn, C., Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net. *Remote Sens-Basel* **2019**, 11, (3), 312.
35. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C., Deep learning – Method overview and review of use for fruit detection and yield estimation. *Comput Electron Agr* **2019**, 162, 219-234.
36. Li, W.; Dong, R.; Fu, H.; Yu, L., Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks. *Remote Sens-Basel* **2018**, 11, 11.
37. Li, W.; Fu, H.; Yu, L.; Cracknell, A., Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sens-Basel* **2017**, 9, (1), 22.
38. Santos, A.A.D.; Marcato Junior, J.; Araújo, M.S.; Di Martini, D.R.; Tetila, E.C.; Siqueira, H.L.; Aoki, C.; Eltner, A.; Matsubara, E.T.; Pistori, H.; Feitosa, R.Q.; Liesenberg, V.; Gonçalves, W.N., Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. *Sensors* **2019**, 19, (16), 3595.
39. Safonova, A.; Guirado, E.; Maglinets, Y.; Alcaraz-Segura, D.; Tabik, S., Olive Tree Biovolume from UAV Multi-Resolution Image Segmentation with Mask R-CNN. *Sensors-Basel* **2021**, 21, (5), 1617.
40. 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-Basel* **2022**, 14, (2), 295.
41. Everingham, M.; Van Gool, L.; Williams, C.; Winn, J.; Zisserman, A., The Pascal Visual Object Classes (VOC) challenge. *Int J Comput Vision* **2010**, 88, (2), 303-338.
42. Wang, C.; Liao, H.M.; Wu, Y.; Chen, P.; Hsieh, J.; Yeh, I., In *CSPNet: A New Backbone that can Enhance Learning Capability of CNN*, Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 14-19 June 2020, 2020; Seattle, WA, USA, **2020**; pp. 1571-1580.
43. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J., In *Path Aggregation Network for Instance Segmentation*, Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018, 2018; Salt Lake City, UT, USA, **2018**; pp. 8759-8768.
44. He, K.; Zhang, X.; Ren, S.; Sun, J., Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Ieee T Pattern Anal* **2015**, 37, (9), 1904-1916.