Computers and Electronics in Agriculture 199 (2022) 107191

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A novel approach for the 3D localization of branch picking points based on deep learning applied to longan harvesting UAVs

Denghui Li [a,b,1], Xiaoxuan Sun [c,d,1], Shengping Lv [a], Hamza Elkhouchlaa [a], Yuhang Jia [a], Zhongwei Yao [a], Peiyi Lin [a], Haobo Zhou [a], Zhengqi Zhou [a], Jiaxing Shen [e,*], Jun Li [a,b,*]

[a] College of Engineering, South China Agricultural University, Guangzhou 510642, China
[b] Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510640, China
[c] Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[d] University of Chinese Academy of Sciences, Beijing 100049, China
[e] Department of Computing, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Keywords:
Picking drones
Convolutional Neural Network
Image analysis
Three-dimensional localization

ABSTRACT

Longan is a famous speciality fruit and cultivated medicinal plant that has important edible and medicinal value; how to improve productivity in harvest is an important issue. At present, longan is mainly planted in hilly areas. For complex site conditions and tall trees, the ground harvesting machineries cannot work normally. In this study, aiming at harvesting longan fruit using unmanned aerial vehicles, a method combining an improved YOLOv5s, improved DeepLabv3+ model and depth image information is proposed, which is used for the three-dimensional (3D) positioning of branch picking points in complex natural environments. First, the improved YOLOv5s model is used to quickly detect longan fruit skewers and the main fruit branches from a complex orchard environment. The correct main fruit branch is obtained according to its relative position relationship and is extracted as the input to the semantic segmentation model. Second, using the improved DeepLabv3+ model, the image extracted in the previous step is semantically segmented to obtain the 2D coordinate information of the main longan fruit branches. Finally, combined with the growth characteristics of a longan fruit string, RGB-D information fusion is carried out on the main fruit branches in 3D space to obtain the central axis and pose information of the main fruit branches, and the 3D coordinates of the picking points are calculated, which provides destination information for a longan harvesting drone. To verify the effectiveness of the proposed method, an experiment for identifying and locating the main fruit branches and picking points was carried out in a longan orchard. The experimental results show that the longan string fruit and main fruit branch detection accuracy is 85.50%, and the main fruit branch semantic segmentation accuracy is 94.52%. The whole algorithm takes 0.58 s in the actual scene and can quickly and accurately locate the picking points. In summary, this paper fully exploits the advantages of the combination of a convolutional neural network and RGB-D image information, further improving the efficiency of longan harvesting drones in accurately positioning picking points in 3D space.

1. Introduction

Longan is a famous specialty fruit and cultivated medicinal plant in tropical and subtropical areas, that has important edible and medicinal value and is widely planted in hilly areas of southern China (Lin et al., 2020). However, longan trees are usually more than 10 m high and have a short maturity period, which requires considerable labour and high costs to harvest. At present, longan harvesting mainly adopts manual operation, which is low in automation, laborious and time-consuming. It is easy for longan pulp quality to deteriorate due to not harvested in time. Therefore, to reduce the harvesting cost of longan, it is necessary to develop an agricultural robot that can automatically harvest longan string fruits (Kang & Chen, 2020; Li et al., 2020; Zhang et al., 2020).

* Corresponding authors at: College of Engineering, South China Agricultural University, Guangzhou 510642, China. E-mail addresses: jiaxshen@polyu.edu.hk (JX. Shen), autojunli@scau.edu.cn (J. Li).
E-mail addresses: jiaxshen@polyu.edu.hk (J. Shen), autojunli@scau.edu.cn (J. Li).
[1] Denghui Li and Xiaoxuan Sun are the co-first authors.

https://doi.org/10.1016/j.compag.2022.107191
Received 1 April 2022; Received in revised form 27 June 2022; Accepted 30 June 2022
Available online 13 July 2022

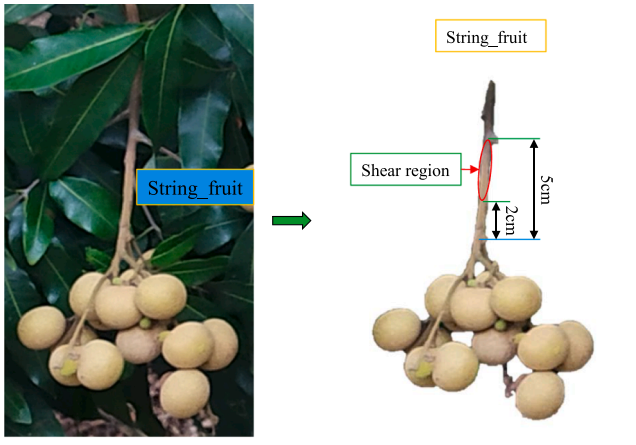
Nomenclature

UAV	Unmanned Aerial Vehicle
CNN	Convolutional Neural Network
RGB-D	Red, Green, Blue and Depth
YOLO	You Only Look Once
3D	Three Dimensional
HSB	Hue, Saturation and Brightness
YOLACT	You Only Look At Coefficien Ts
R-CNN	Recursive Convolutional Neural Network
ASPP	Atrous Spatial Pyramid Pooling
LoG	Laplacian of Gaussian
FPN	Feature Pyramid Network
SPP	Spatial Pyramid Pooling
PAN	Path Aggregation Network
P	Precision
R	Recall
AP	Average Precision of a category
mAP	Average Precision of multiple categories
FPS	Frames Per Second
IoU	Intersection over Union
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
P-R	Precision-Recall
PA	Pixel Accuracy
mPA	mean Pixel Accuracy
mIoU	mean Intersection over Union

Especially considering the complex terrain conditions of mountain orchards and the growth characteristics of fruit clusters on tall longan trees, it is necessary to develop more suitable harvesting robots.

The autumn tip of longan trees usually grows at the top of the periphery, which is an important fruiting branch. After flowering and fruit inoculation, with the increase in fruit weight, it gradually bends towards the ground (Pham et al., 2015). Longan, like litchi and grape, is a cluster fruit and the whole cluster needs to be picked. When picking longan fruit manually, it is necessary to find the main branch first, which is then cut with scissors at a distance of 2–5 cm from the last branch to prevent the fruit from being damaged. The growth characteristics and picking scheme of cluster fruit are shown in Fig. 1. For a machine to automatically pick longan, it is necessary to imitate manual harvesting, first, accurately identifying and locating the main fruit branches and then finding the picking point. In a natural environment, the main longan fruit branches have complex distribution forms, which are easily shaded by leaves and branches and show different postures in different growing environments. Therefore, detecting the main longan fruit branches and the positioning results of picking points are difficulties in realizing automatic picking, which directly affects the accuracy and efficiency of longan picking.

Some researchers have used traditional machine learning methods to identify string fruits. Jaisin et al. (2013) applied a HSB (hue, saturation and brightness) colour model to the images of longan fruits in bunches, including branches and leaves, to separate the objects of interest from the background. Xiong et al. (2018) proposed a method based on improved fuzzy clustering to separate litchi fruit and the main fruit-bearing branches and calculate the three-dimensional (3D) coordinates of picking points with binocular vision. Zhuang et al. (2019) used the retinex algorithm to segment litchi's main fruit bearing branch area and then used Harris corner points to determine the picking point. The colour space transformation, fuzzy clustering, threshold segmentation and other methods used in the above research are traditional image processing technologies. The image data used in the research are all collected under simple background conditions or indoor conditions and can only be used for image processing tasks with simple backgrounds. In an actual orchard scene, due to the influence of various factors in the natural environment, the fruit growing environment varies greatly, and the light intensity also varies with the weather conditions. The image data collected in these scenes not only have complicated backgrounds but also images, where light and dark often alternate, and the characteristics of different fruits are obviously different. Therefore, these algorithms have poor robustness and accuracy in complex orchard scenes.



Growth characteristics of string fruit Schematic diagram of suitable cutting area

Fig. 1. Growth characteristics and picking scheme of string fruits.

With their rapid development, deep convolutional neural networks have shown excellent learning ability in the feature extraction of complex images. Therefore, an increasing number of deep learning (DL) algorithms are used to process image data collected in the agricultural field (da Silva et al., 2021; de Medeiros et al., 2021; Kamilaris & Prenafeta-Boldu, 2018), including fruit detection and counting (Bargoti & Underwood, 2016; Fu et al., 2020; Gao et al., 2020; Liu et al., 2020; Xiong et al., 2020), plant identification (Dyrmann et al., 2017; Flores et al., 2021), pest identification and diagnosis (Anagnostis et al., 2021; Ghosal et al., 2018; Singh et al., 2021), remote sensing area classification and detection (Ma et al., 2019; Paoletti et al., 2019), fruit in vivo detection and product classification (Koirala et al., 2019), and animal identification and posture detection (Norouzzadeh et al., 2018). Compared with traditional image processing algorithms, the multilayer structure of a DL model forms abstract high-level representation attribute categories or features by combining bottom features. It can solve complex nonlinear problems well, has strong robustness in agricultural applications and shows high performance in object recognition (LeCun et al., 2015).

To realize the automatic picking of string fruits, researchers have attempted to apply DL methods to detecting and locating string fruits in recent years. Liang et al. (2020) used YOLOv3 and U-Net (Ronneberger et al., 2015) to detect and segment litchi fruit and the main fruit-bearing branches but did not further determine the actual picking point. Zhong et al. (2021) proposed a detection method based on YOLACT, which determines the angle of the main fruit-bearing branch in two-dimensional space through skeleton extraction and least squares fitting but did not propose a calculation method in 3D space. Li et al. (2021) proposed a scheme based on YOLOv4-MobileNet to quickly and accurately detect and locate suitable picking points of longan branches but did not further determine the position and posture information of

the main longan branches.

With their rapid development, unmanned aerial vehicles (UAVs) have been gradually applied in agricultural production, including plant protection (Tetila et al., 2020; Zhou et al., 2020), crop monitoring (Feng et al., 2020a; Vanegas et al., 2018) and crop yield evaluation (Feng et al., 2020b; Sumesh et al., 2021). Compared with ground harvesting robots, UAVs are more adaptable to the complex terrain of mountain orchards. Therefore, this study attempts to apply UAVs to picking tasks in unstructured orchard environments. With the development of sensor integration, RGB-D cameras are more suitable for outdoor orchard environments because of their lightness, high precision and insensitivity to light, and they have become an effective tool for collecting 3D orchard information. At present, carrying a portable RGB-D camera on a UAV to collect images, accurately detecting the main longan branch and locating the picking point based on DL have become key problems for longan harvesting UAVs using vision to complete the picking task.

To promote the application of UAVs in longan picking, this study proposes a scheme that combines a DL algorithm with an RGB-D camera to accurately detect and locate the main longan branch and determine the picking point. This scheme will help longan harvesting UAVs using vision improve the speed and accuracy of object location in natural environments. The contribution of this research can be summarized as follows:

(A) At present, the general object detection model requires considerable computation and running time. In this study, a dense cross-connection method is used to improve the YOLOv5s network performance so that the model can converge to the optimal solution faster to obtain more target information and improve the accuracy of the model in the object detection task.

(B) Directly using the semantic segmentation model to segment small targets in a large field of view images consumes considerable running time, and the accuracy is not high. This study proposes using object detection before semantic segmentation, which fully improves the performance of the whole algorithm, reduces the amount of calculation, parameters and running time of the model, and further improves the accuracy and efficiency of target positioning.

(C) Given the large error of binocular vision in target positioning, a method fusing a convolutional neural network and depth information is proposed in this study to estimate the pose of longan main fruit branches and accurately locate the picking points in three-dimensional space, which is expected to improve the robustness of UAVs to accurately perceive targets under complex and changing conditions.

(D) The method of collecting RGB and depth images with RGB-D cameras on UAVs proposed in this study will help provide a large quantity of data for UAVs in any field to perform accurate object detection and semantic segmentation tasks.

The article is organized as follows: the materials and methods are presented in Section 2, the model construction and 3D localization strategy are described in Section 3, the model experiment and results analysis are detailed in Section 4, and Section 5 concludes the paper.

2. Materials and methods

2.1. Overview of the 3D localization system

To improve the speed and accuracy of locating the main longan fruit branches and picking points in 3D space, a system solution was developed, as shown in Fig. 2. The scheme provides a UAV image acquisition method and image preprocessing method, a scheme integrating target detection, a semantic segmentation model and depth image information for 3D positioning of longan main fruit branches and picking points in a complex natural environment.

During image acquisition, the RGB-D camera on the UAV captures RGB and depth images from a distance of 1.5 m to 0.6 m from the longan fruit. The improved YOLOv5s object detection model is used to detect longan skewers and main fruit branches in a complex natural environment. The correct main fruit branch detection frame is obtained according to its relative position relationship, and it is extracted as the input to the semantic segmentation model. An improved DeepLabv3 + model is used to perform semantic segmentation on the image extracted in the previous step and obtain the 2D coordinate information of the longan main fruit branches. Finally, combined with the growth characteristics of the longan fruit string, RGB-D information fusion is carried out on the main fruit branches in 3D space, and the pose information of the main fruit branches and the coordinate information of the picking points are determined, which provides destination information for the longan harvesting drone.

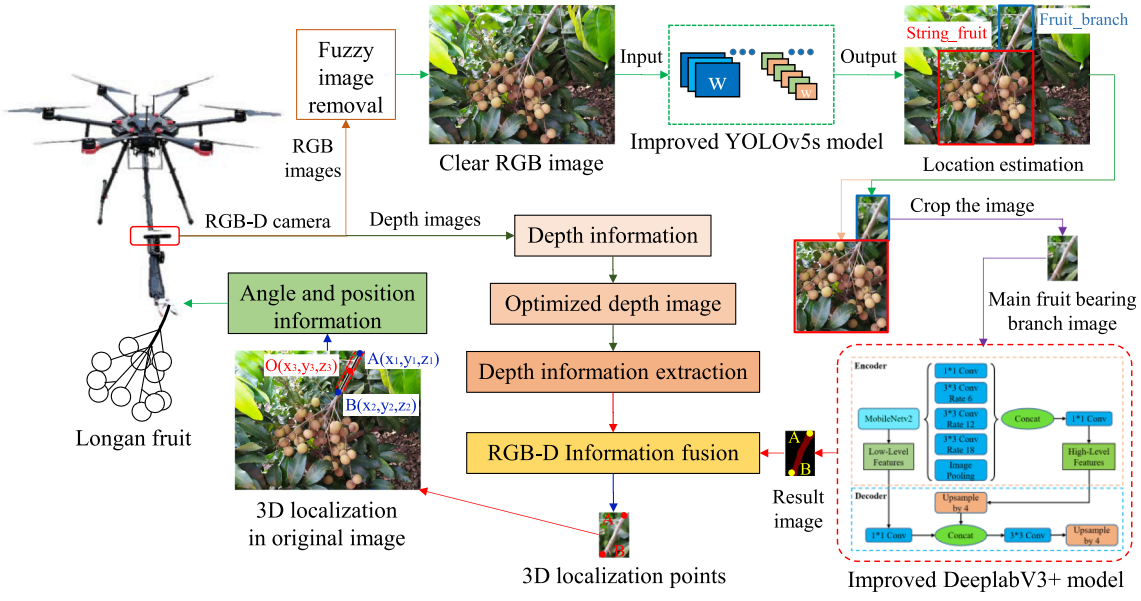


Fig. 2. 3D positioning solution of the main longan branch.

2.2. Sensor system and image acquisition

To develop and test the proposed algorithm, 450 and 360 valid longan images were acquired at the Longan Germplasm Resource Nursery of the Guangdong Academy of Agricultural Sciences in Guangzhou on July 1–17, 2020, and July 5–25, 2021, respectively, during three different time periods: morning (8:00–10:00), noon (12:00–14:00), and afternoon (15:00–17:00). An Intel RealSense D455, a lightweight, small and powerful RGB-D camera, was installed on a DJI Jingwei M600PRO UAV. The camera consisted of a colour camera and an infrared camera. The installation method and structure are shown in Fig. 3, and the resolution of the output RGB image and depth image was set to 1280 × 720 pixels.

To enhance the generalization of the model, images of Shijie and Chuliang longan were collected. To fully reflect the real scene and the complexity of an orchard environment when picking longan strings by UAV, images collected in different scenes, such as sunny, backlit and cloudy in the natural state, were used as data sets, without artificial shadows or light interference. Examples of the images are shown in Fig. 4. The onboard computer consisted of two ARM v8 64-bit CPUs, one 8 GB 128-bit LPDDR4 memory, one NVIDIA Pascal GPU architecture with 256 NVIDIA CUDA cores, and CUDA version 9.0. An Ubuntu 16.04 LTS 64-bit system, based on Jetpack SDK 3.3, a PyTorch DL framework was built, and the image acquisition and data processing program was written in Python. The CNN model program was written on a Python platform to realize the object detection and semantic segmentation tasks. The data processing of the whole platform runs on an airborne microcomputer.

2.3. Image preprocessing

This study prepares training data and test data for target detection and semantic segmentation models respectively. Aiming at the RGB images collected by the UAV, blurred images are identified and cleared, the clear images are randomly cut and normalized, and the border boxes of the string fruit and the main fruit branches in each sample are manually marked. An initial data set for training and testing the improved YOLOv5s model is constructed. The trained improved YOLOv5s model is used to detect every image in the clear RGB image data set, and according to the position relationship between the string fruit and the main fruit branch, the correct main fruit branch image is extracted, and the position of the main fruit branch in each image is manually marked. Semantic segmentation of the data set is performed for training and testing the improved DeepLabv3 + model. The specific flow of the entire process is shown in Fig. 5.

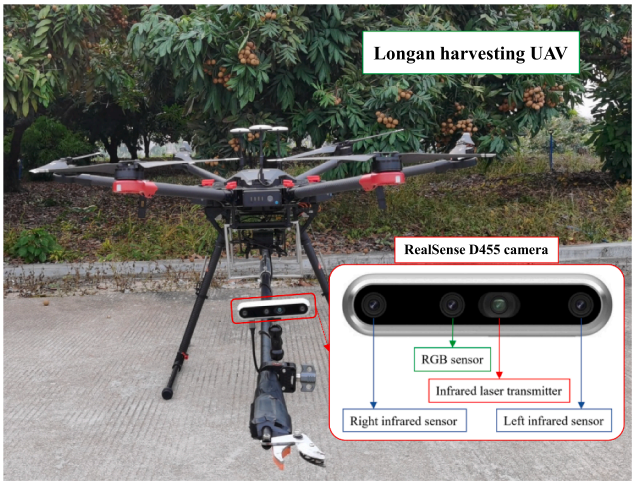


Fig. 3. Installation method and structure of the Intel RealSense D455 camera.

2.3.1. Fuzzy image recognition

When collecting longan images in hilly orchards with complex terrain, a UAV needs to adjust its posture at any time according to the terrain changes and the shape of the fruit trees. At the same time, it is easily affected by local circulation and airflow in orchards, which will cause unstable posture in the air. Therefore, the collected longan images will inevitably be blurred. To improve the accuracy of object detection and positioning for a longan harvesting UAV, it is necessary to judge whether the collected images are clear in real time. In the early research of this project, the method of extracting edge features was used to judge the blurred image and remove the blurred image (Li et al., 2021). This method is used to process 910 collected images, 125 images are identified and removed, and the remaining 785 images are processed in the next step.

2.3.2. Construction and annotation of the object detection data set

In the process of building the target detection data set, 785 clear RGB images obtained in the previous step were expanded and normalized. To ensure the diversity of training samples, the longan image data were expanded by using the self-programming random clipping algorithm. A total of 1,070 longan images were obtained after the expansion of the initial dataset. To prevent the image size inconsistency from adversely affecting the training process of the object detection model, image normalization technology was used to preprocess the amplified image, and the size of 1280 × 1280 pixel standard image was obtained.

In images collected by a UAV, the background is complicated. Because the string fruit is easier to accurately identify than the main fruit branch using the model, it is necessary to further judge whether the identified main fruit branch is correct according to the position information of the string fruit detected by the model. Each of the 1,070 images was manually annotated. To draw the bounding boxes, we followed the guidelines of the reference challenge Pascal VOC 2010. Two classification labels were defined: ① string_fruit: a cluster of longan fruit from the first branch to the last branch on a fruiting parent branch; ② fruit_branch: the main fruit branch.

LableImg software was used for manual annotation, and the annotation information was saved in an XML file. Labelling information included image size, object category and specific location coordinate information of the object area, which was used as an information file to read object features during model training. For example, as shown in the image in the second column in Fig. 6, for the four scenes of A, B, C and D, string_fruit and fruit_branch are marked according to the following situations: ① when string_fruit and fruit_branch on the same fruit branch are both visible or partially occluded, marking the two kinds of objects at this time can calculate the reverse loss of model training, optimize the model parameters and evaluate the performance of the model; ② when string_fruit and fruit_branch on the same branch cannot be seen at the same time, it cannot be determined whether the positional relationship between string_fruit and fruit_branch is accurate, and the end effector cannot pick fruit, so it is not necessary to mark string_fruit and fruit_branch in this case.

2.3.3. Construction and annotation of the semantic segmentation data set

Since the semantic segmentation model used in this study does not require the size of the input image, the trained improved YOLOv5s model is used to detect each image in the initial data set, obtain the coordinate information of the main fruit branch image, and extract it directly from the original image. Lableme software is used to manually mark the position of the main fruit branches in each image to form a semantic segmentation data set for training and testing the improved DeepLabv3 + model.

The sample and tag information of the semantic segmentation dataset are shown in the third column of the main fruit branch image and the fourth column of the semantic labels in Fig. 6. The semantic segmentation model is trained, and its performance is evaluated through the tag mask and tag. The initial dataset and the semantic segmentation

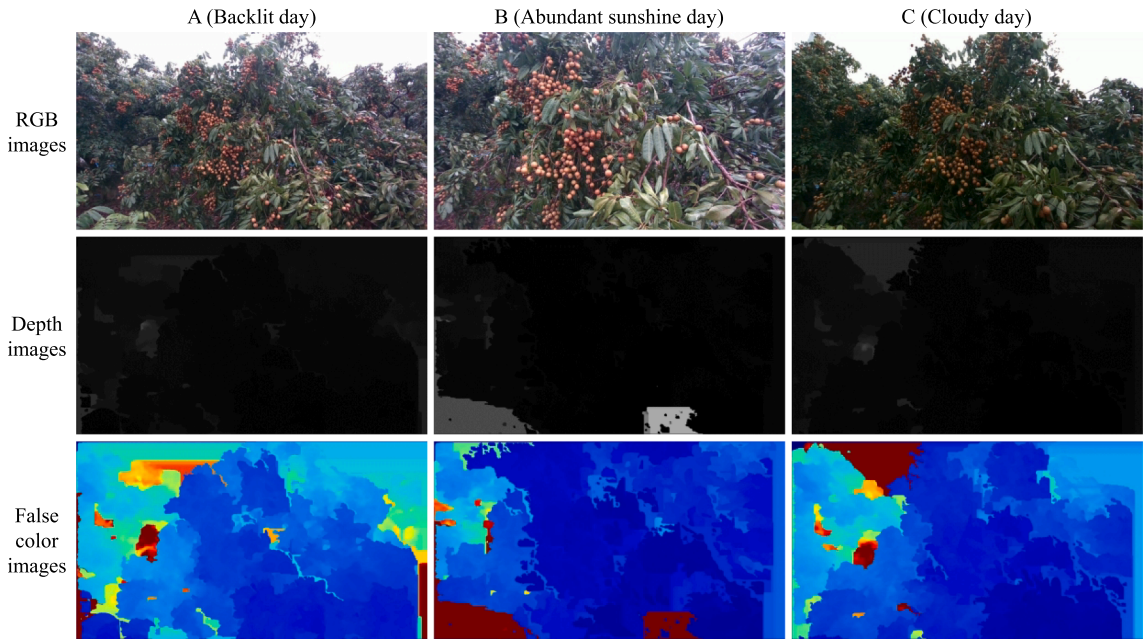


Fig. 4. Examples of longan images.

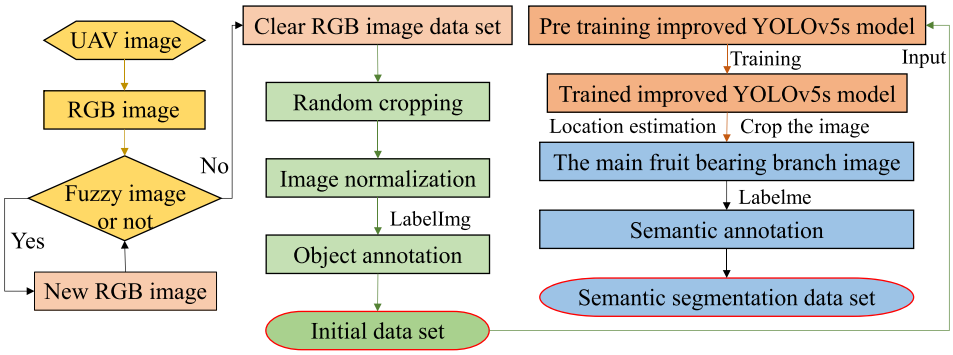


Fig. 5. Diagram of the UAV image preprocessing flow.

dataset can be obtained from the RGB images collected by the UAV after image processing. The 1,070 images in the initial data set are divided into a training set, verification set and test set according to the ratio of 3:1:1. Since the semantic segmentation model used in this study only needs a training set and test set, 818 images in the semantic segmentation data set are divided into a training set and test set according to the ratio of 4:1. Table 1 lists the number of images and labelling information contained in the two datasets.

3. Model construction and 3D localization strategy

In this section, focusing on RGB and depth images collected by an RGB-D camera on a UAV, we propose a strategy for quickly and accurately obtaining the position information of string_fruit and fruit_branch from RGB images, combining this with depth images, extracting the pose information of longan main fruit branches in 3D space and accurately positioning picking points. The specific scheme is shown in Fig. 7. First, RGB images are input to the trained improved YOLOv5s model for object detection, and the model outputs the coordinate information of string_fruit and fruit_branch in the RGB images. According to their relative position, the correct image of the main fruit branch is judged and extracted from the original image. This is input into the improved DeepLabv3 + model for semantic segmentation. The location information of the main fruit branch in the original image is obtained, and the information in the depth image is fused to obtain the 3D coordinate information of the centroid of the local areas at both ends of the main fruit branch. Finally, according to the 3D space angle calculation method, the angle information between the central axis of the main fruit branch and the XOY, YOZ and XOZ planes is obtained, and the positioning information of the picking point in 3D space is obtained. In the following four parts, the object detection algorithm, semantic segmentation model, judgement strategy of the main fruit branch position and precise positioning strategy of the picking points are introduced.

3.1. YOLOv5 object detection algorithm

To quickly and accurately detect the longan location and further reduce the number of calculations and parameters, it is necessary to optimize a CNN model suitable for deployment on the onboard UAV computer. At present, the main object detection networks include the R-CNN series (Girshick, 2015; Girshick et al., 2016; Ren et al., 2017) and YOLO series (Bochkovskiy et al., 2020; Redmon et al., 2016; Redmon & Farhadi, 2018). The R-CNN series has advantages in object detection with high precision, but in practical application scenarios, it cannot meet real-time requirements. YOLO series algorithms use the idea of regression and single-stage neural networks to directly detect and classify objects, which makes it easier to learn the generalized features of objects, thus improving the object detection speed.

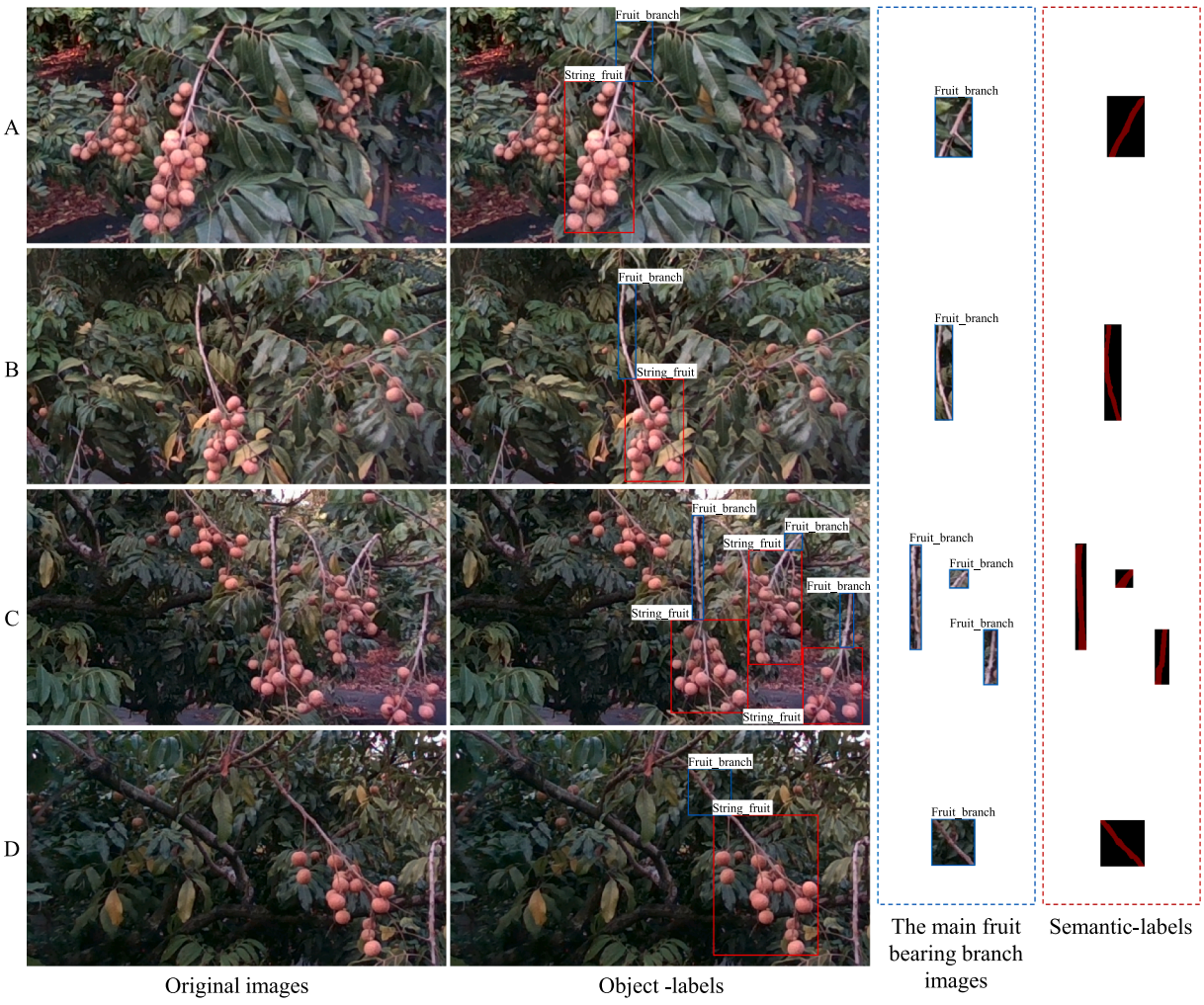


Fig. 6. Construction and annotation of the initial and semantic segmentation data sets.

Table 1
Details of the initial and semantic segmentation data sets.

Initial data set			Semantic segmentation data set		
Data set	Images	string_fruit	fruit_branch	Images	fruit_branch
All data set	1070	1573	1573	818	818
Train data set	642	950	950	654	654
Val data set	214	305	305	/	/
Test data set	214	318	318	164	164

YOLOv5s can be structurally divided into four parts: the input, backbone, neck and prediction. Its network structure is shown in Fig. 8. The input terminal is used to process the image to some extent. The backbone is used to downsample the input image five times to extract the features of the image. It includes four modules: Focus, CBH, CSP1_x and SPP (He et al., 2015). The neck network uses a ReLU activation function and adopts an FPN (Lin et al., 2017) + PAN (Liu et al., 2018) network structure. In the prediction, NMS performs nonmaximum suppression processing on the last detection frame of the object to obtain the optimal object frame and provides three different detection scales (20 × 20, 40 × 40 and 80 × 80). It can predict different sizes of longan fruit clusters and the main fruit branches.

It can be seen in the structural diagram of YOLOv5s that there is one CSP1_1 and two CSP1_3 modules in the backbone, and these three residual components are connected by cross-connection, which easily leads to model gradient divergence in the training process, thus leading to the problem that the accuracy is not easy to improve. To further improve the detection accuracy of the model, this study uses a dense cross-connection to improve the performance of the YOLOv5s network. The structural improvement method is shown in Fig. 9.

The improved CSP1_x residual component adds a cross-connection between each CBH module based on the original residual component so that the two CBH modules are closely connected. The specific process is as follows: (1) the input feature information (defined as A) in the first CBH module is convolved and the obtained feature information (defined as B) is fused with the original input information; (2) then the fused features in the second CBH module are convolved to obtain feature information (defined as C); and (3) the feature information A, B and C are fused and input into the next feature extraction structure. The improved CSP1_x residual component can prevent the gradient degradation of the YOLOv5s model during training so that the improved YOLOv5s model can converge to the optimal solution faster, thereby obtaining more object information and improving the accuracy of the model in the object detection task.

3.2. *DeepLabv3 + semantic segmentation algorithm*

When a CNN is deployed on a UAV's onboard computer to realize semantic segmentation tasks, it is necessary to reduce the number of calculations and parameters while ensuring high accuracy. Facing the problem of continuous pooling and downsampling of images, most semantic segmentation algorithms adopt a method of atrous convolution, but this method cannot solve the problem of multiscale objects. The

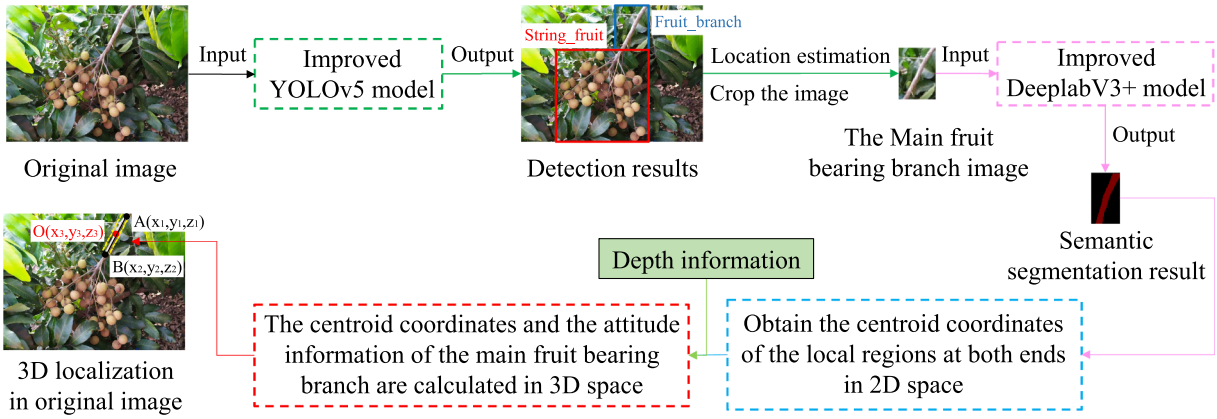


Fig. 7. 3D positioning strategy of the longan picking point.

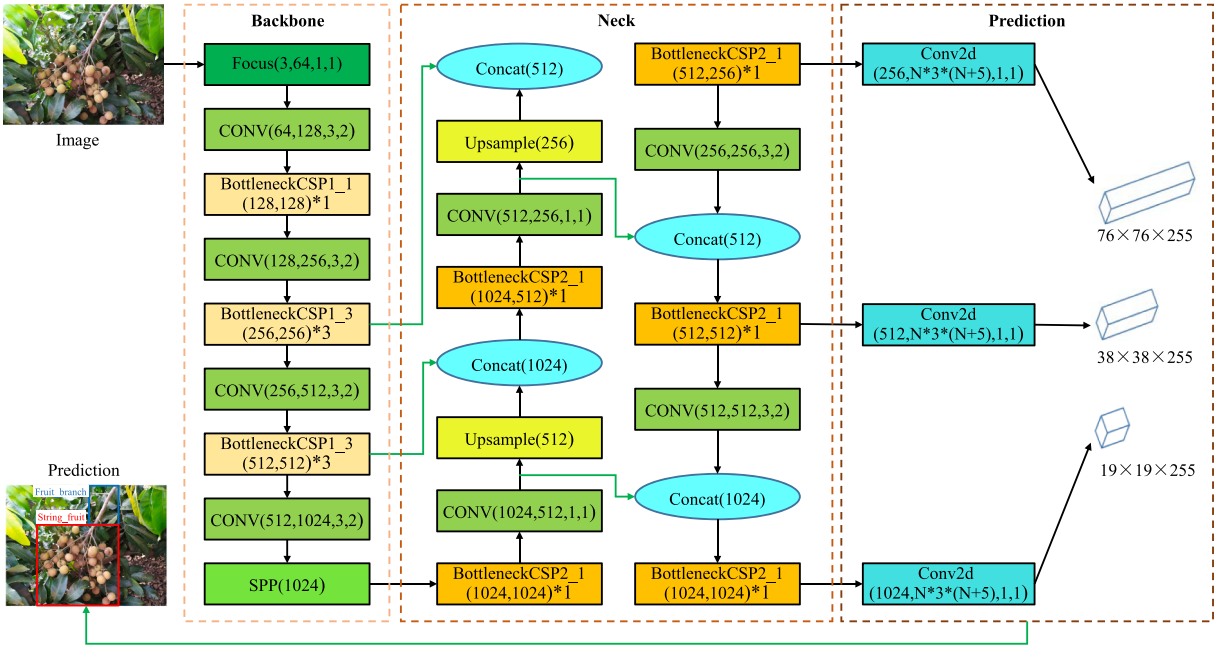


Fig. 8. YOLOv5s structure diagram.

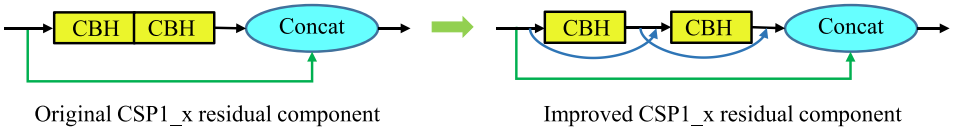


Fig. 9. CSP1_x residual component before and after improvement.

DeepLabv3 + algorithm was developed based on DeepLabv1-3 and other algorithms (Baheti et al., 2020; Chen et al., 2018; Chen et al., 2017).

To further improve the speed and accuracy of the DeepLabv3 + model, this study uses MobileNetv2 as the feature extraction backbone network and optimizes the existing DeepLabv3 + model. This can not only greatly reduce the number of model parameters and realize light-weight model design but also ensure that deep convolution can complete feature extraction in high dimensions and improve the calculation performance of the model. The improved network structure is shown in Fig. 10. It is divided into an encoding layer and a decoding layer.

In the encoding layer, the MobileNetv2 network and ASPP module are used to extract object features. The ASPP structure first performs 1 × 1 convolution, 3 × 3 convolution with expansion rates of 6, 12 and 18, and global average pooling operation on the input feature map, then performs feature fusion on the feature information generated after parallel convolution, inputs the feature information into the 1 × 1 convolution layer for compression, and outputs advanced semantic feature information. In this process, after the 1 × 1 convolution, the number of channels is compressed to 256, which is consistent with the low-level semantic information of the decoding layer. Finally, ASPP can extract and distinguish the feature information of objects of different scales and realize the segmentation of multiscale objects well.

In the decoding layer, the low-level semantic information generated by the encoding layer is reduced in dimension by a 1 × 1 convolution, which makes the latter features have a strong emphasis on the features obtained by the encoder, that is, the 256 channels, and well preserves

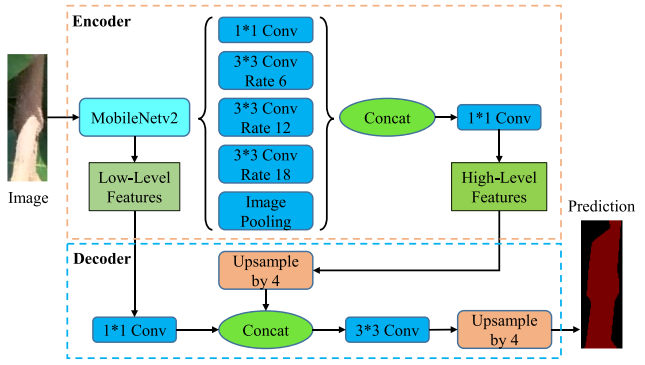


Fig. 10. Network structure diagram of the improved DeepLabv3+ model.

the high-level feature information to compensate for the loss of boundary information caused by downsampling. Low-level features contain rich resolution and rich spatial details, and feature fusion is carried out with the advanced semantic information after upsampling 4 times. Finally, 3 × 3 convolution and 4 upsampling operations are carried out on the fused feature map to restore the spatial resolution of the image, and the object segmentation result is output after passing through a softmax layer.

3.3. Judgement strategy of the main fruit branch position

In this section, it is determined whether the position of the main longan branch that is output is accurate according to the object detection result chart. Usually, longan string fruit is vertical to the ground because of its heavy branches, and the main fruit branch is above the string fruit. Because of the large area of radiation from the fruit skewers, it is easier to accurately detect the specific location in both a distant view and close view. Therefore, it can be judged whether the detected main fruit branch is the real main fruit branch according to the position of the fruit string.

As shown in Fig. 11, in the result chart predicted from the improved YOLOv5s model, the red box is the string fruit position of the model detection output, and the blue box is the main fruit branch position of the output. The coordinate information of the detection frame predicted by the model is calculated by taking the upper left corner of the image as the origin, and the unit is pixels. The numerical values of the coordinates of the upper left corner point and the lower right corner point of each detection frame are sequentially output from left to right and from top to bottom. To accurately judge whether the main fruit branch predicted by the model is the correct main fruit branch, a verification strategy is set: $Xs1 < Xf1 < Xs2$ and $Xs1 < Xf2 < Xs2$ and $Yf1 < Ys1$ and $Yf2 < Ys2$. According to this verification strategy, it can be easily judged that only blue box 1 in Fig. 11 is the correct main fruit branch, while neither blue box 2 nor blue box 3 is the correct main fruit branch, so only blue box 1

Fig. 11. Coordinate information for each bounding box.

needs to be extracted as the semantic segmentation image.

3.4. Accurate picking point location strategy

Longan fruit is usually harvested in clusters. When picking, it is necessary to first obtain the position information and picking point of the main longan branches. Under normal circumstances, the fruit stringing is under the main fruit branch, and the upper part of the main fruit branch is covered by leaves. The centre of the main fruit branch is selected as the picking point to avoid the end actuator touching the fruit or leaves when cutting the main fruit branch, thereby improving the picking success rate.

Two pieces of information for the pose information of the main fruit branches and the location of picking points need to be acquired first, the two-dimensional coordinate information from the RGB image and the distance information from the depth image. The grey value of each pixel in the depth image is linearly related to the distance between the point and the camera, and the specific distance between the object and the camera can be extracted by picking up the grey value of a pixel at a specific position. The pose of the main fruit branch is defined using three angles, which are the angles between the main fruit branch and the XOY, YOZ and XOZ planes in 3D space. The picking point is defined as the central position of the central axis of the main fruit branch. Their specific calculation steps are described in steps 1–6. The steps of manually measuring the position and posture information of the main fruit branch and the position information of the picking point are described in steps 7–8.

(1) Step 1: As shown in Fig. 12, after the RGB image collected by the camera is predicted by the object detection algorithm, the coordinate points of the upper left corner and the lower right corner of the bounding box of the main fruit branch in the RGB image can be obtained as (X_{min}, Y_{min}) and (X_{max}, Y_{max}), respectively. According to these two coordinates, the RGB image of the main fruit branch is extracted from the original image.

(2) Step 2: After the RGB image obtained in the previous step is predicted by the semantic segmentation model, the segmentation result map of the main fruit branch in the RGB image is obtained, the result map is binarized so that the area with a pixel value of 255 is the main fruit branch, and the area with a pixel value of 0 is the background.

(3) Step 3: The upper and lower end regions of the main fruit branch obtained by semantic segmentation are usually irregular. To accurately obtain the pose information of the main fruit branch, the local regions of the upper and lower ends of the region with a pixel value of 255 are selected, and the coordinates of their centroids A (X_{centA}, Y_{centA}) and B (X_{centB}, Y_{centB}) are calculated. Then, the coordinates of centroids A and B of the local areas at the upper and lower ends of the main fruit branch in the two-dimensional space are $(X_{min} + X_{centA}, Y_{min} + Y_{centA})$ and $(X_{min} + X_{centB}, Y_{min} + Y_{centB})$.

(4) Step 4: As shown in Fig. 13, the numerical values in the parameter matrix of the Intel RealSense D455 camera f_x, f_y, pp_x, pp_y can be read directly, and the centroid coordinates in two-dimensional space (X_p, Y_p) and depth values z can be obtained. After that, RGB-D information can be fused using the pinhole imaging principle, and the 3D coordinates of the centroid points in the real world can be obtained (X_t, Y_t, Z_t):

$$\begin{cases} X_t = \dfrac{z^*(X_p - pp_x)}{f_x} \\ Y_t = \dfrac{z^*(Y_p - pp_y)}{f_y} \\ Z_t = z \end{cases} \tag{1}$$

According to this formula, the coordinate values of centroids A and B

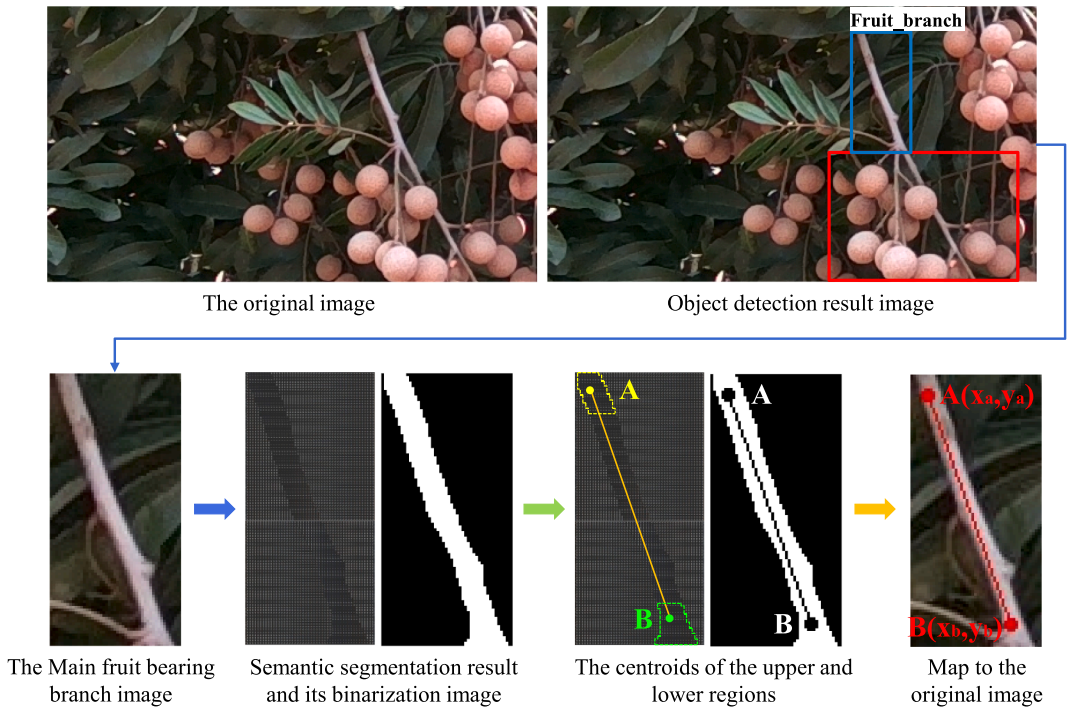


Fig. 12. Flow chart of obtaining coordinates of the centroid in two-dimensional space.

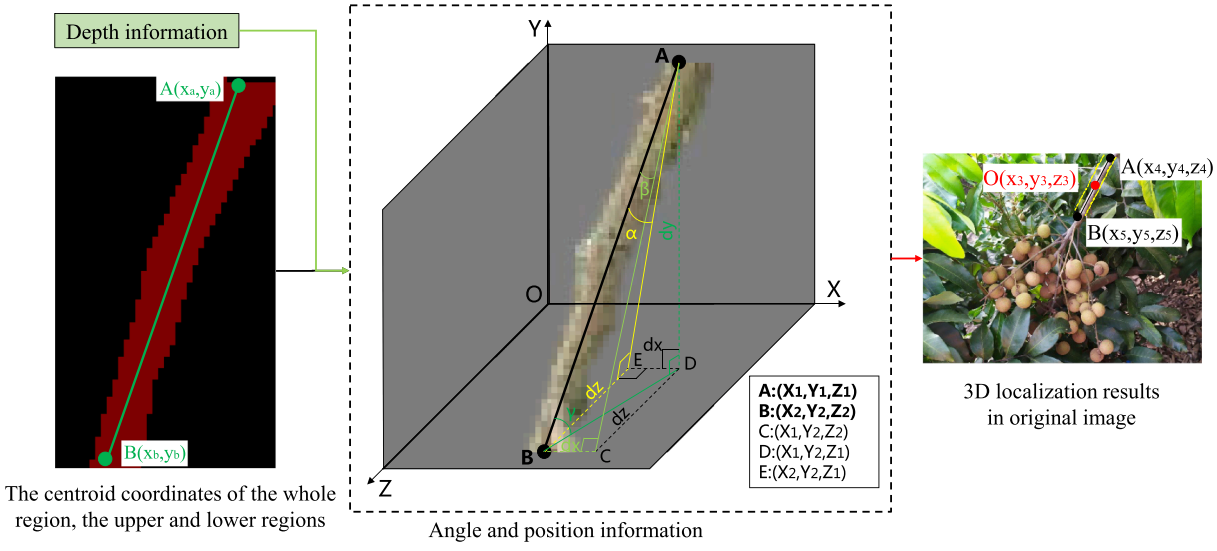


Fig. 13. Flow chart of obtaining the picking point coordinates in 3D space.

in 3D space can be calculated to be (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2), respectively. The line segment between points A and B is used to fit the central axis of the main fruit branch.

(5) Step 5: As shown in Fig. 13, according to the 3D coordinates of points A and B, the angle between the main fruit branch and each plane can be calculated. When calculating, $dx = |X_2 - X_1|$, $dy = |Y_2 - Y_1|$, $dz = |Z_2 - Z_1|$ are first defined. The pose information of the main fruit branch is obtained as follows:

Angle between the main fruit branch and the XOY plane:

$$\alpha = \arctan[dz/sqrt(dx*dx + dy*dy)] \quad (2)$$

Angle between the main fruit branch and the YOZ plane:

$$\beta = \arctan[dx/sqrt(dy*dy + dz*dz)] \quad (3)$$

Angle between the main fruit branch and the XOZ plane:

$$\gamma = \arctan[dy/sqrt(dx*dx + dz*dz)] \quad (4)$$

(6) Step 6: According to the 3D spatial coordinates of points A and B and the fitted central axis of the main fruit branch, the coordinates of the central point O (X_3, Y_3, Z_3) are calculated, where point O is the best picking point on the main longan branch.

(7) Step 7: When manually measuring the α, β, γ with the help of a herringbone ladder, tripod with platform, level, white paper, square and two in one angle ruler. In the preparation stage of measurement, the researcher stood on a herringbone ladder and

placed the level near the lower end of the main fruit branch on the platform supported by the tripod. A piece of white paper was posted on the platform. The platform was adjusted to be in the horizontal position by the level. With the help of a square, the projection of the upper and lower points of the appropriate cutting area on the main fruit branch on the horizontal plane was drawn on white paper on the platform. Additionally, the BCDE rectangle in Fig. 13 was drawn on the white paper and the line segment between the two points B and D was drawn as the auxiliary measuring line of the two in one angle ruler. The measure is as follows:

(a) Close the rotation centre of the two in one angle ruler to point A in Fig. 13. Stick the bottom edge of the ruler with a digital display on the angle ruler to the AE line and the side edge to the BE line. Turn the other ruler on the angle ruler until its bottom edge is attached to the AB line. Read the angle value on the digital display as the α value.

(b) Close the rotation centre of the two in one angle ruler to point An in Fig. 13. Stick the bottom edge of the ruler with a digital display on the angle ruler to the AC line and the side edge to the BC line. Turn the other ruler on the angle ruler until its bottom edge is attached to the AB line. Read the angle value on the digital display as the β value.

(c) Close the rotation centre of the two in one angle ruler to point B in Fig. 13. Stick the bottom edge of the ruler with a digital display on the angle ruler to the BD line and the side edge to the AD line. Turn the other ruler on the angle ruler until its bottom edge is attached to the BA line. Read the angle value on the digital display is the γ value.

(8) Step 8: When manually measuring the coordinates of the best picking point on the main fruit branch in 3D space, first use a ruler to measure the distance between two points A and B in Fig. 13, and calculate the centre point O of the two points A and B. Use a square to obtain the intersection M between the plane where point O is located and the vertical line where the camera centre point is located, and use a ruler to measure the distance Z_4 between point M and the camera centre point. Use a ruler to measure the distance X_4 between point M and point O in the X-axis direction. Distance Y_4 between point M and point O in the Y-axis direction. The coordinates of the best picking point on the main fruit branch in 3D space are (X_4, Y_4, Z_4).

4. Model experiment and results analysis

The focus of this section analyses the performance of the models in combination with the experimental results. The following will shows the model training and parameter design, model evaluation indicators, the results and discussion of the object detection task, the best segmentation result of the main fruit branch, and the results of the extraction and location of picking points in real orchard scenes.

4.1. Model training and parameter design

4.1.1. Object detection algorithm training and parameter design

The training and testing process of the object detection model is realized on a workstation with the Ubuntu 16.04 LTS operating system. The CNN model is built using the Python programming language on the Python DL framework, and the image size of the training data is 1280 × 1280 pixels. In terms of parameter settings, the intersection over union (IoU) is set to 0.5, the initial learning rate is 0.0001, the learning rate at the end of training is set to 0.00001, and 90–10 training verification data set splitting is used. A total of 300 epoch iterative trainings are conducted. To enable the model to detect multiple objects at multiple scales in one window, it is necessary to set anchor boxes with different aspect ratios and sizes. By analysing the initial data set, the size distribution map and k-means clustering result map of all object bounding boxes are obtained. In the left figure of Fig. 14, the red rectangular boxes correspond to the string_fruit bounding boxes in the dataset. The blue rectangular boxes correspond to the fruit_branch bounding boxes in the dataset. As seen in the figure, the size of fruit_branch is generally small. The horizontal and vertical coordinates of the blue dots on the right figure of Fig. 14 represent the width and height of each object bounding box. The darker the colour, the more objects of this size there are. In other words, this map can reflect the complexity of an object to be inspected in an orchard scene to a certain extent.

According to the size characteristics of the object bounding boxes in the initial dataset, it is determined that the anchor boxes suitable for training the improved YOLOv5s model are [6,14, 8,33, 13,20, 17,43, 35,60, 59,88, 89,137, 148,214, 293,339].

4.1.2. Semantic segmentation algorithm training and parameter design

The training and testing hardware configuration of the semantic segmentation model is as follows: Ubuntu 16.04 LTS is the operating system, an Intel Core i7-10700 k @ 3.00 GHz is the CPU, there are 512 GB of memory, and an NVIDIA GTX2070 Super is the GPU. The semantic segmentation model is built using the Python programming language on the PyTorch DL framework. In addition, the semantic segmentation model runs on the GPU configured with the CUDA 10.0 parallel programming platform and the cuDNN 7.1 acceleration package. To avoid the possible influence of different parameter settings on the experimental results, the optimal parameters are obtained through repeated experiments, and then the parameters of each network are configured consistently. In the freeze phase, the learning rate is 5e-4, and 200 epochs are trained. In the unfreeze phase, the learning rate is 5e-5, and 300 epochs are trained. The number of iterations of each epoch is 164, and the batch size is 4.

4.2. Model evaluation indicators

4.2.1. Evaluation index of the object detection algorithm

The commonly used algorithm evaluation indicators in object detection tasks include the precision (P), recall (R), F1 score, average precision of a category (AP), average precision of multiple categories (mAP), and frames per second (FPS). For binary classification problems, samples can be divided into true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) according to the combination of their true categories and the learner prediction categories. In four cases, let TP, FP, TN, and FN denote the corresponding number of samples. Obviously, TP + FP + TN + FN = the total number of samples.

The precision, which means how many of the positive examples are true examples, is based on the prediction results. The recall, which indicates how many positive examples in the sample are predicted to be correct, refers to the original sample. The F1 score indicates the harmonic mean of P and R when λ = 0.5. The methods for calculating the precision, recall and F1 score are shown in formulas (5), (6) and (7).

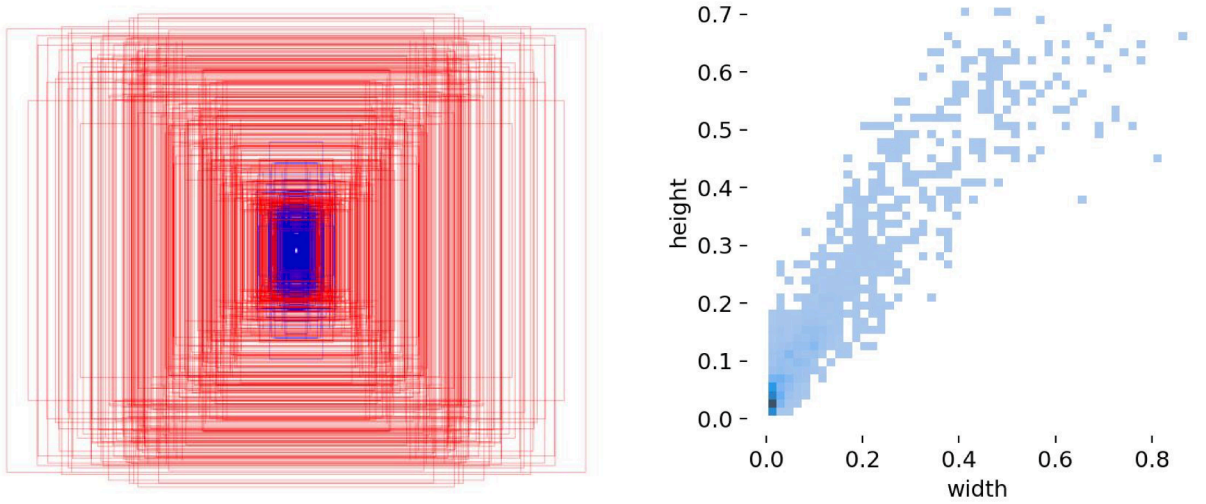
$$\mathrm{P} = \frac{TP}{TP + FP} \tag{5}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{F1\ score} = \frac{2{*}P{*}R}{P + R} \tag{7}$$

In the formula, TP, FP, TN, and FN represent true cases, false positive cases, true negative cases, and false negative cases, respectively.

The AP and mAP are comprehensive evaluation indicators that are proposed to solve the single-point value limitation of indicators such as the precision and recall. They are indicators that can reflect the global performance of the model. AP is the average accuracy rate, and is the integral of the P index to the R index, that is, the area under the P–R curve. After calculating the AP value, the average value of all APs can be



Size distribution diagram of bounding box

K-means clustering result diagram of bounding box

Fig. 14. Visualization result diagram of object bounding box features in the initial data set.

obtained to obtain the mAP value. They are defined as follows:

$$\mathrm{AP} = \int_0^1 P(R)d_R \tag{8}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP(i) \tag{9}$$

where N is the number of categories.

4.2.2. Evaluation index of the semantic segmentation algorithm

The indexes commonly used to measure the accuracy of algorithms in image semantic segmentation include the pixel accuracy (PA), mean pixel accuracy (mPA) and mean intersection over union (mIoU). For convenience of explanation, assume the following: k +1 classes (from L_0 to L_k, which contain an empty class or background), p_{ij} represents the number of pixels that belong to class i but are predicted as class j; that is, p_{ii} represents the real quantity, and p_{ij} p_{ji} represent a false positive and false negative, respectively.

The pixel accuracy is the proportion of correctly marked pixels to total pixels; that is, the main longan branch obtained by model segmentation is compared pixel-by-pixel with the standard main longan branch obtained by manual labelling and the PA is calculated. The mean pixel accuracy is the proportion of correctly classified pixels in each class, and then the average of all classes is calculated. The mean intersection over union is used to calculate the ratio of the intersection and union of two sets, which are the ground truth and the predicted segmentation from semantic segmentation. This ratio can be transformed into the sum of true positives over true positives, false negatives and false positives. IoU is calculated for each class, and then all classes are averaged. The F1 score indicates the harmonic mean of P and R when λ = 0.5. The F1 score formula is shown in (7). The PA, mPA and mIoU formulas are shown in (10), (11) and (12), respectively:

$$\mathrm{PA} = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \tag{10}$$

$$\mathrm{mPA} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{11}$$

$$\mathrm{mIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{12}$$

where k is the total number of categories (k +1 when including the background), p_{ij} represents the total number of i pixels that are predicted as j pixels, and p_{ii} represents the total number of i pixels that are predicted as i pixels. Since the first stage aims to segment leaves from the background, k = 1.

4.3. Results and discussion of the object detection task

4.3.1. Comparison of object detection task results

To fully evaluate the performance of the improved YOLOv5s model, first, the identification performance of the four versions of the YOLOv5 model and the improved YOLOv5s model is compared and analysed using the evaluation index in Section 4.2.1. By default, the same training parameters are set for the five models, and the files with the best training effect in each model are saved as weight files and then used for testing. When testing the five models, 82 test images randomly divided from the initial data set are input into the five models, and the detection result maps of each model are obtained. Finally, the models are comprehensively evaluated based on the AP, mAP, FPS, F1 score, weight file size, training time, etc., and the object detection model that is most suitable for the UAV onboard processor is selected.

The P-R curves and F1 score changes of the five models on the same test dataset for the detection of two types of objects, string_fruit and fruit_branch, are shown in Fig. 15. The area of the area enclosed by the P-R curve and the two coordinate axes is the AP value of the corresponding classification. According to Table 2, the AP values of the five models on the test set for string_fruit are all above 86%, and the P-R curve basically covers the entire coordinate system. The F1 score changes slightly with the increase in the confidence value at first and suddenly decreases sharply when the confidence value is greater than 0.85. Therefore, it is usually sufficient to set this parameter to 0.5 in the model training stage. According to the distribution of object size in the dataset, the size of fruit_branch is generally small. However, the F1 score of the five models on string_fruit is only slightly larger than that on fruit_branch. This shows that YOLOv5 also has good detection performance for small objects.

The comparison results are shown in Table 2. The data in the table shows that the improved YOLOv5s has a better recognition result than other networks, except that the recognition accuracy of fruit_branch is lower than that of YOLOv5l. In particular, compared with the original YOLOv5s, the recognition accuracy of string_fruit is 7.5% higher, reaching 94%, and the map is 5.3% higher. The improved YOLOv5s only takes 17 ms to detect an image on average, and its detection speed is

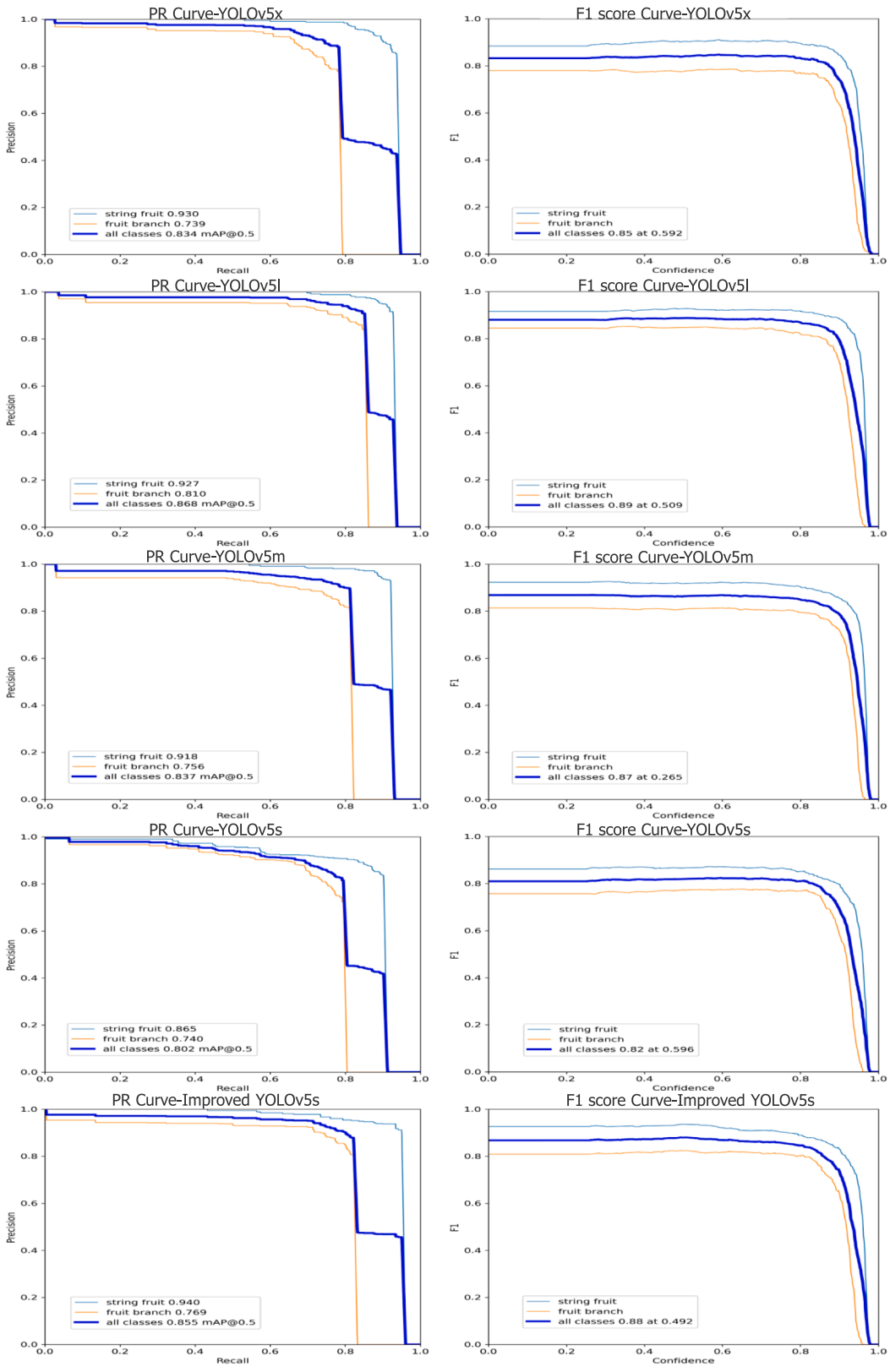


Fig. 15. P-R curves and F1 score for different detection methods.

more than twice that of YOLOv5m, more than four times that of YOLOv5l and more than eight times that of YOLOv5x. It can greatly improve the detection efficiency of longan harvesting by UAVs and meet the needs of real-time detection.

The F1 score of the improved YOLOv5s is slightly lower than that of YOLOv5l, but its weight file size is only one-seventh of that of YOLOv5l. The improved YOLOv5s is used to identify longan fruit, and it can greatly reduce the number of calculations and parameters in the

Table 2
Evaluation index results of the test set using different YOLOv5 models.

Model	AP (%)		mAP (%)	FPS	F1 score	Weight file size (MB)	Training time (h)
	string_fruit	fruit_branch					
YOLOv5x	93.00	73.9	83.40	7.36	0.85	175.20	14.903
YOLOv5l	92.7	81.00	86.8	13.33	0.89	94.00	8.983
YOLOv5m	91.8	75.6	83.7	23.81	0.87	42.70	6.103
YOLOv5s	86.50	74.00	80.20	55.62	0.82	14.60	3.205
Improved YOLOv5s	94.00	76.90	85.50	58.82	0.88	14.60	3.205

detection process and further reduce the energy consumption of the microcomputer carried on the UAV. In addition, when training 300 epochs on the training data set, the training time of the improved YOLOv5s model is only one quarter of that of YOLOv5x, which is beneficial to saving time in the training phase of the model. In summary, in this study, the improved YOLOv5s is deployed to the UAV airborne microcomputer to perform the object detection task.

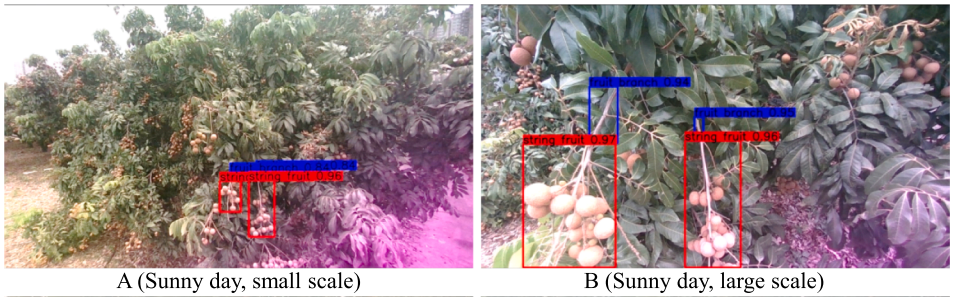
4.3.2. Detection effects of the improved YOLOv5s in real scenes and on different varieties

To further evaluate the performance of the improved YOLOv5s model in detecting longan stringing and main fruit branches in a real and complicated mountain orchard environment, this section selects longan orchard images with different illumination, different scales and different densities and tests the trained improved YOLOv5s model. Four examples in each environment are selected for illustration. Fig. 16A and B show the detection results under sunny conditions. Whether on a large or small scale, detection is accurate. Fig. 16C and D show the detection results under cloudy conditions. Whether in sunny or cloudy weather, string_fruit and fruit_branch are accurately detected. It is not difficult to see from the figure that each object in an image is accurately detected under different illumination conditions and different scales. The detection results show that the improved YOLOv5s model has good feature extraction performance, is not easily disturbed by uneven light in a real orchard environment, has strong generalization, and has a good detection effect on small objects. It is suitable for object detection in longan picking.

There are many varieties of longan, and new varieties have appeared in recent years. To verify that the improved YOLOv5s model can realize robust detection of different longan varieties in real orchards, especially different varieties and sizes of longan in the same scene, Fig. 17A and B show the detection results of Chuliang longan and Fig. 17C and D show the detection results of Shixia longan. There are great differences in fruit colour and fruit shape and size between the two kinds of longan, which show different lustres at different distances and under different illumination. It can be seen in the test result chart that both string_fruit and fruit_branch are accurately detected. This shows that the improved YOLOv5s model has strong generalization to different varieties of longan.

4.4. The best segmentation result of the main fruit branch

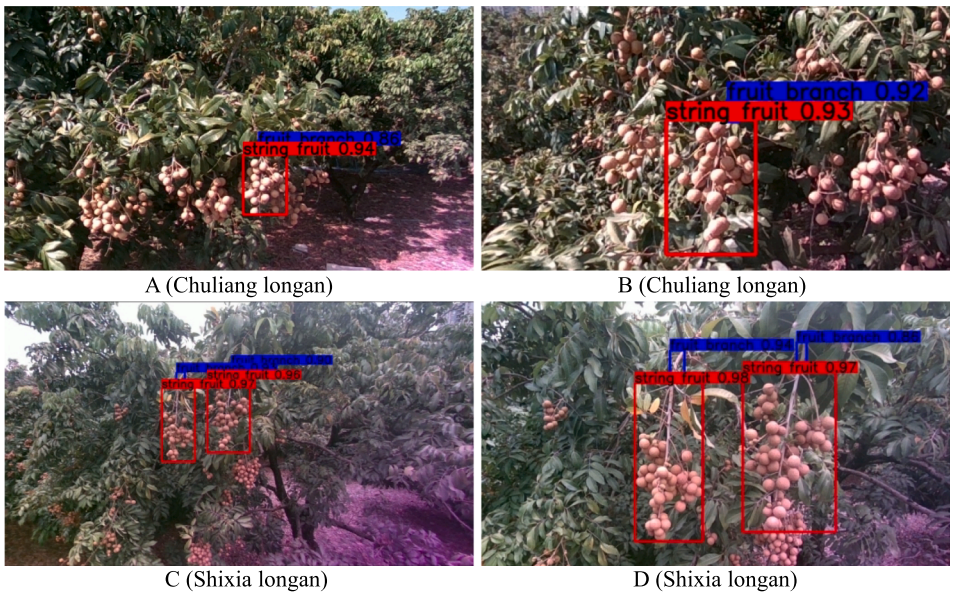
Because the bearing weight of longan harvesting UAVs is limited, it is necessary to further reduce the number of calculations, number of parameters and energy consumption of the airborne processor while ensuring object positioning accuracy. Therefore, aiming at the task of semantic segmentation of main fruit branches, this study proposes to improve DeepLabv3 + and adopts MobileNetv2 as the backbone network of the DeepLabv3 + model for feature extraction. To verify the performance of the improved model in semantic segmentation of the main fruit branches, the experimental results are compared with the PSPNet (MobileNet and ResNet-50 are used as the backbone networks for feature extraction), UNet and DeepLabv3 + before improving classical semantic segmentation networks. The five models are comprehensively evaluated from the aspects of the mPA, mIoU, FPS, weight file size, and F1 score. The comparison results are shown in Table 3. The results of semantic segmentation of the main fruit branches by the five models in real orchard scenes are shown in Fig. 18. The red area is the main fruit branch, and the black area is the background. A, B, C and D in the figure are the results of semantic segmentation of the main fruit



A (Sunny day, small scale) B (Sunny day, large scale)

C (Cloudy day, small scale) D (Cloudy day, large scale)

Fig. 16. Detection results under different scenes.



A (Chuliang longan) B (Chuliang longan)

C (Shixia longan) D (Shixia longan)

Fig. 17. Detection results of different varieties of longan.

Table 3
Evaluation index results on the test set using different semantic segmentation models.

Model	mPA (%)	mIoU (%)	FPS	F1 score	Weight file size (MB)
PSPNet (MobileNet)	74.91	61.36	102.70	0.72	9.25
PSPNet (ResNet-50)	86.70	78.21	39.10	0.79	178.00
UNet	91.02	84.36	18.27	0.81	94.90
DeepLabv3+	91.64	85.11	23.54	0.83	209.00
DeepLabv3+ (MobileNetv2)	94.52	89.69	71.51	0.88	22.40

branches in different light intensity scenes, and the light intensity gradually weakens.

Table 3 and Fig. 18 show that the improved DeepLabv3 + model and the other four models can realize segmentation of the main fruit branches, but there are great differences among the models. Among the recognition results of the PSPNet (MobileNet), PSPNet (ResNet-50) and UNet models, the PSPNet (MobileNet) model has the fastest recognition speed and the smallest weight file size, but its recognition accuracy is the lowest. There is a large gap between the predicted effect picture and the original picture, and there are many recognition errors.

As an ideal model for semantic segmentation of uncooperative fruit branches, the PSPNet (ResNet-50) model has a relatively balanced index, which can realize the identification of main fruit branches, but there are still a few obvious errors, so the main fruit branches in some scenes with weak light are identified as the background. UNet has high recognition accuracy, which can recognize main fruit branches, but its recognition speed is the slowest, and the weight file size is larger. Compared with the DeepLabv3 + model, the improved DeepLabv3 + model has higher recognition accuracy, and the recognition speed is three times that of the original model, but the weight file size is only one-ninth of that of the original model, and the F1 score is the highest. Comparing the prediction map with the annotation map, the improved model can better segment the main fruit branches in different light intensities and scenes. However, the original model has some recognition errors in weak light intensity. Overall, the mPA and mIoU of the improved DeepLabv3 + model are better than those of the other network structures, and it has the best effect on the main fruit branch segmentation, showing good generalization ability, faster feature extraction, smooth edge processing of the prediction map and better detail processing. Therefore, the improved DeepLabv3 + model is incorporated into the airborne processor, not only improving the accuracy and recognition efficiency of longan harvesting UAV object location but also reducing the number of calculations and parameters of the airborne processor and the energy consumption of the UAV, which has strong economic value.

4.5. The results of extraction and picking point locations in real orchard scenes

In a real orchard scene, longan fruit grows on a branch of the main fruit branch, which is a longan cluster composed of several small branches. In the process of longan ripening, as the fruit becomes heavier, eventually the fruit branches will hang downwards in the outer canopy. Based on this growth feature, the advantages of the combination of a CNN and RGB-D camera are fully exploited, and picking point locations in 3D space are obtained. The following aspects are discussed from the best positioning result of the central axis of the main fruit branch, the estimation result of the position and posture of the main fruit branch, the positioning result of the picking point and the experimental results of the full pipeline.

4.5.1. The best location result of the central axis of the main fruit branch

Usually, the main longan branch in an orchard is not a straight cylinder due to the influence of natural external forces, fruit gravity and the mutual interference of branches and leaves. When the camera collects an image of the main fruit branch, the main fruit branch will be partially blocked by other branches and leaves. In summary, the influence of the above factors and the semantic segmentation results of the main fruit branches obtained using semantic segmentation are usually irregular. When determining the centroid of the upper and lower ends of a suitable shearing area on the main fruit branch, to obtain the best positioning result, according to steps 3 and 4 in Section 3.4, when selecting the local areas of the upper and lower ends of the main fruit branch, 5 rows of pixels, 10 rows of pixels, 15 rows of pixels and 20 rows of pixels are selected as local areas at the upper and lower ends from the top and the lowest end, respectively. The corresponding centroid coordinates are calculated, and the two centroids are connected with line segments to compare which line segment has the best fitting effect with the actual axis of the main fruit branch.

According to the above method, the main longan branches in different scenes in an orchard were selected, and the centroids of their

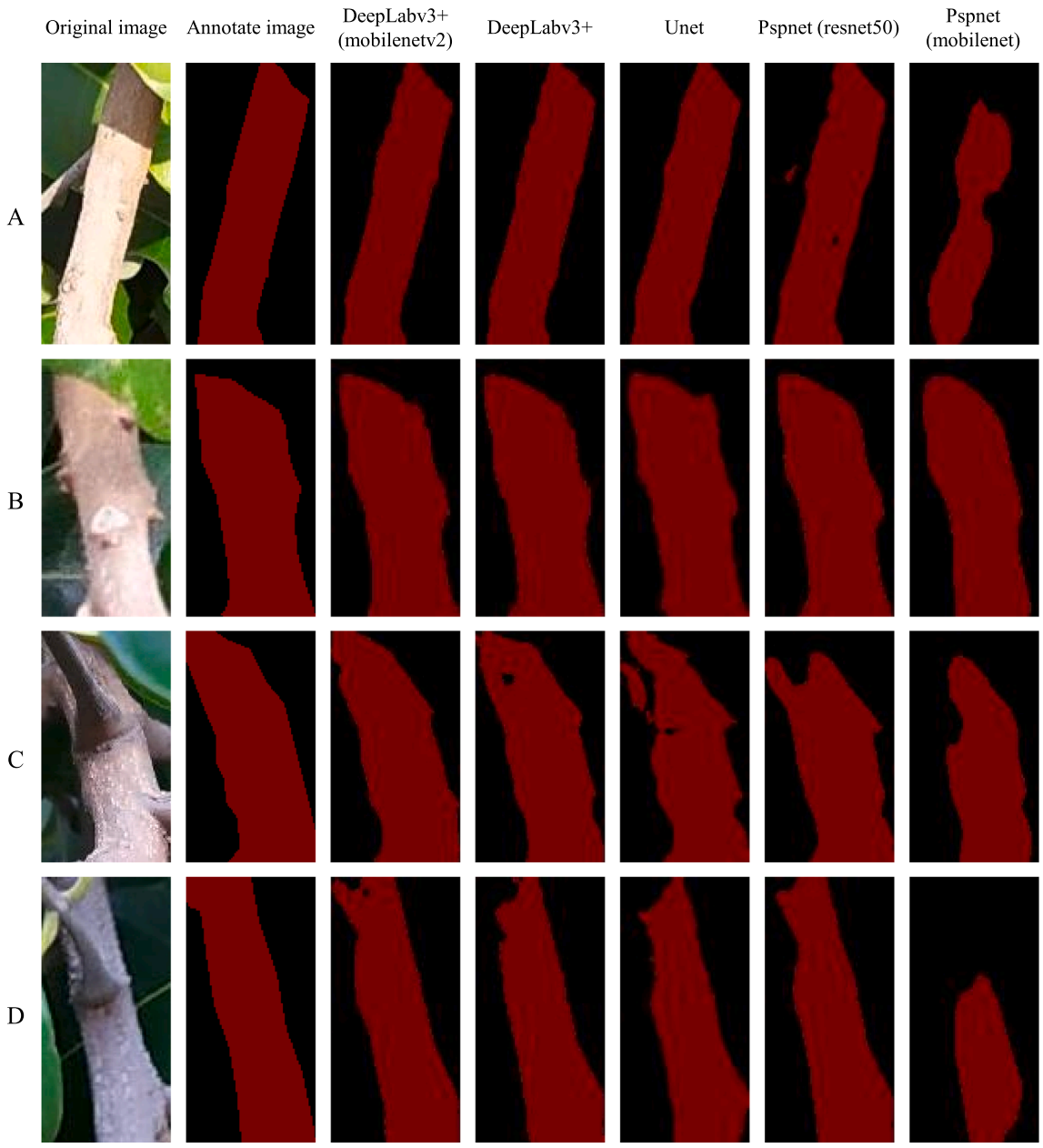


Fig. 18. Semantic segmentation results of the five models in different scenes.

upper and lower ends and the connecting line segments were obtained, which were compared with the true central axes of the main branches. As shown in Fig. 19, A, B, C and D are the main longan fruit branches in different scenes, and the blue line is the central axis of each main fruit branch. The red area is the semantic segmentation area corresponding to the main fruit branch. The centroids of the upper and lower ends obtained through calculation are represented by green dots, and the connection between two green dots is the predicted central axis of the main fruit branch. The obtained centroid and central axis are mapped to the original image, which are represented by red dots and red line segments, respectively. As seen in the figure, with the increase in the number of pixels in the upper and lower end areas, the error between the true central axis of each main fruit branch and the predicted axis decreases. In the B and C scenes, the two axes completely coincide. When 20 rows of pixels are selected as the local areas at the upper and lower ends in the A and D scenes, the error between them is also very small. Therefore, when predicting the central axis of the main fruit branch, 20 rows of pixels are selected as local areas at the upper and lower ends of the main fruit branch, and the best positioning result of the main fruit branch can be obtained.

4.5.2. *Position and pose estimation results of the main fruit branches*

The estimation result of the position and posture of the main fruit branch directly determines the picking angle and the accurate positioning result of the picking point. Based on the centroid coordinates of the local areas at the upper and lower ends of the main fruit branch and the central axis of the main fruit branch obtained in the previous section, the included angles α, β and γ between the main fruit branch and the XOY, YOZ and XOZ planes can be calculated according to step 5 in Section 3.4. To verify the effectiveness of this research method, 10 main fruit branches were randomly selected from real orchards and their predicted values of α, β, and γ. At the same time, according to the step 7 method in Section 3.4, the researchers calculated the actual measured values of α, β, and γ. The specific statistical results and their error data are shown in Fig. 20.

As seen in Fig. 20, angles α and β between the main fruit branches

Fig. 19. The location results of the central axis of the main fruit branch in different scenes.

and XOY and YOZ planes are all below 22°, and some are below 10°, while the angles γ between the main fruit branches and XOZ planes are basically above 63°. This is because longan string fruit is subject to gravity, showing growth characteristics towards the ground. There are two main reasons for the error between the measured value and the predicted value. On the one hand, there is natural wind interference in an orchard, which causes the main fruit branches to swing slightly. On the other hand, the distance information between the RGB-D camera and the main fruit branch is obtained by reading the depth image, and the depth image is calculated using the phase difference between the infrared transmitter and receiver of the RGB-D camera, which may be disturbed by the natural environment in an orchard. It can be seen from the data in the error column in the figure that the error between the measured value and the predicted value is generally small, basically within 1.5°. This shows that the angle between the main fruit branches and the XOY, YOZ and XOZ planes predicted using this research method is relatively accurate, which can provide accurate data for the pose adjustment of longan harvesting UAVs.

4.5.3. *Location results of picking points*

The location result of the picking point directly determines the path planning of the longan harvesting UAV and the efficiency of the picking operation. According to the best method determined in Section 4.5.1 and step 6 in Section 3.4, the coordinate information of the picking point can be obtained. To verify the effectiveness of this research method, sixteen longan main fruit branches in real orchards with different lighting conditions, different shading environments and different sizes of main fruit branches were selected. The semantic segmentation result diagram of the main fruit branches was obtained using the improved semantic segmentation method. Twenty rows of pixels were selected as local areas at the upper and lower ends of the main fruit branches, and their centroid coordinates were obtained. Then, the central axis and picking point coordinates of each main fruit branch were calculated.

According to the above method, the location results of the centroid, central axis and picking point of local areas at the upper and lower ends of each main fruit branch are shown in Fig. 21. The red area in the image is the semantic segmentation result of each main fruit branch, and the adjacent image is its original image. In the semantic segmentation result image, the green points at the upper and lower ends are the centroids of

The error between the predicted and measured values of α

Different scenes	Error (°)
Case1	0.7
Case2	1.4
Case3	1.1
Case4	0.8
Case5	1.4
Case6	1.1
Case7	0.5
Case8	0.6
Case9	0.8
Case10	1.1

The error between the predicted and measured values of β

Different scenes	Error (°)
Case1	0.8
Case2	0.7
Case3	0.8
Case4	0.5
Case5	0.6
Case6	0.8
Case7	0.4
Case8	0.6
Case9	0.8
Case10	0.9

The error between the predicted and measured values of γ

Different scenes	Error (°)
Case1	0.7
Case2	0.8
Case3	0.7
Case4	0.4
Case5	0.4
Case6	0.7
Case7	0.4
Case8	0.5
Case9	0.5
Case10	0.4

Fig. 20. Statistical results of α, β, γ and error in different scenes.

the local areas at the upper and lower ends of the main fruit branch, the green line segment is the central axis of the main fruit branch, and the green point in the middle is the picking point. Red dots and lines are used to map the above information to the original image. It can be seen in the figure that the location results of picking points are very accurate in any scene. This shows that the research method proposed in this paper can provide accurate picking point coordinates for longan harvesting UAVs.

4.5.4. Experimental results of the full pipeline

To further evaluate the performance of the full pipeline in the actual harvest scenario, relevant experiments were carried out in longan orchards. The specific process is shown in Fig. 22. Select the scene where the RGB-D camera is 1.5 m to 0.6 m away from the longan fruit, collect RGB images and depth images, and obtain clear RGB images and depth images after removing the blurred images. This process takes 0.16 s.

In each scene, the improved YOLOv5s model is used to obtain the pixel coordinates of the string fruit and the main fruit branch on the RGB image, and the image area of the main fruit branch is extracted

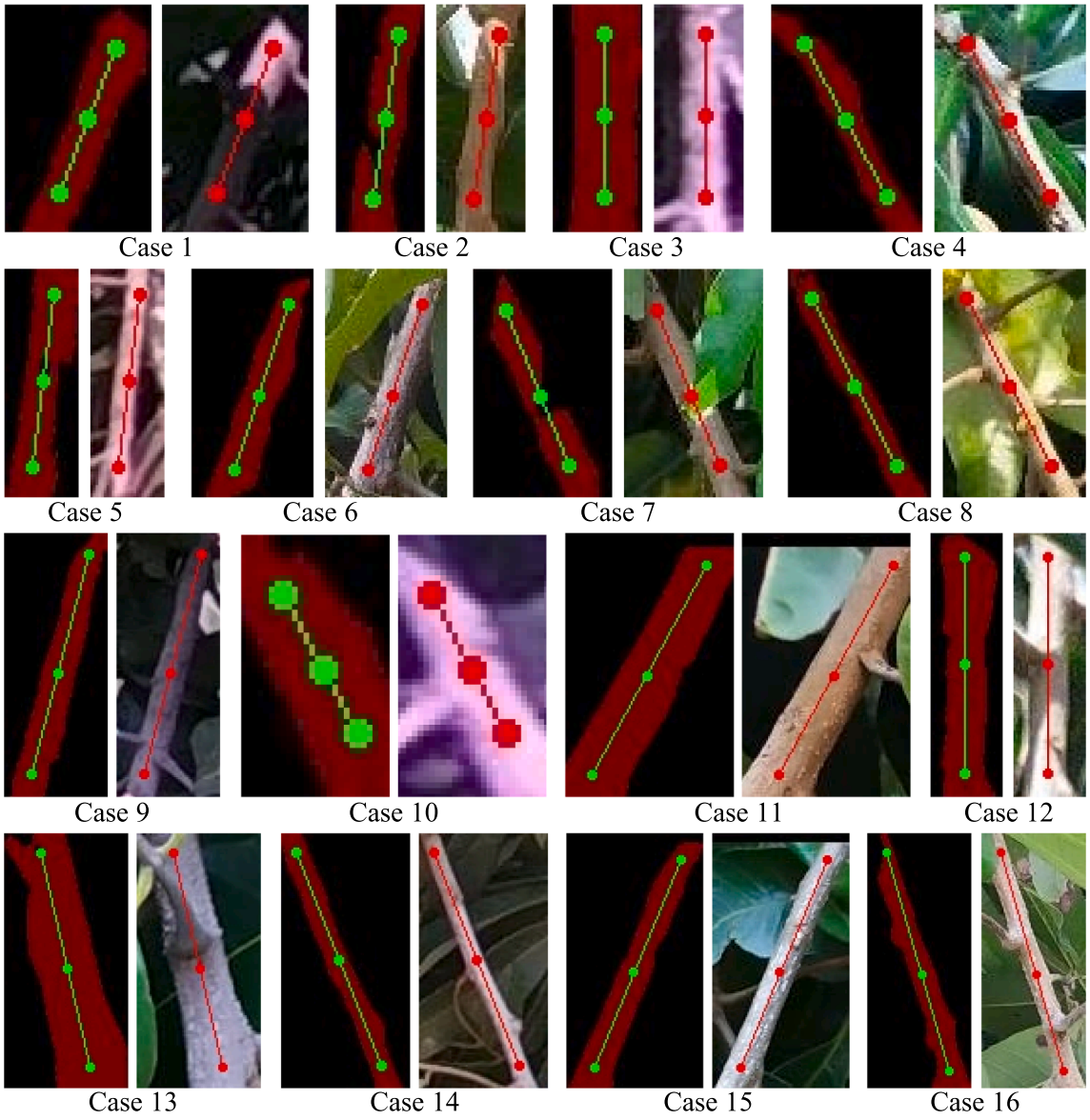


Fig. 21. The location results of picking points in different scenes.

according to the pixel coordinates. This process takes 0.13 s. After the main fruit branch image is input into the improved DeepLabv3 + model, the semantic segmentation result is obtained, and the result is mapped to the original image. This process takes 0.08 s. According to the Step 3 and Step 4 methods in Section 3.4, the centroid coordinates of the local areas at the upper and lower ends of the main fruit branch are fused with the depth information to obtain the coordinates of the centroid in the 3D space. This process takes 0.15 s. According to the step 5 and step 6 methods in Section 3.4, the position and posture information of the main fruit branch and the 3D spatial coordinates of the picking point are obtained. This process takes 0.06 s.

Additionally, according to the step 8 method in Section 3.4, the researchers counted the actual measured values of the mass centres of the upper and lower ends of the main fruit branches and the coordinates of the picking points in 3D space in different distance scenes. The measurement results of the researcher and the positioning results predicted by the whole algorithm and their error data are shown in Table 4.

According to the statistical data in Fig. 22 and Table 4, the accuracy of object detection is above 0.9 in all scenarios, and the accuracy of semantic segmentation is generally high. With the decrease in the distance between the camera and the longan fruit cluster, the proportion of the main fruit branch in the image pixels increases, the accuracy of object detection and semantic segmentation also increases, and the positioning error of the picking point in the 3D space decreases. The error between the researchers' measurement results and the positioning results predicted by the whole algorithm is generally small, which shows that the whole algorithm is effective in positioning the main fruit branches and picking points and can provide an accurate destination for the longan picking UAV. The whole algorithm takes 0.58 s in the actual scene. It can quickly obtain the picking point coordinate information, quickly input the destination to the UAV flight control system, and provide information for flight path planning.

5. Conclusion

In a complex orchard environment, the longan picking point is on the main fruit branch, which is more difficult to locate than a picking point in the geometric centre of the fruit. Therefore, the rapid and accurate positioning scheme of picking point selection proposed in this paper is of great significance and can provide a pair of eyes for fruit harvesting drones. In this paper, a lightweight RGB-D camera is mounted on a longan harvesting UAV and used to collect images of longan orchards.

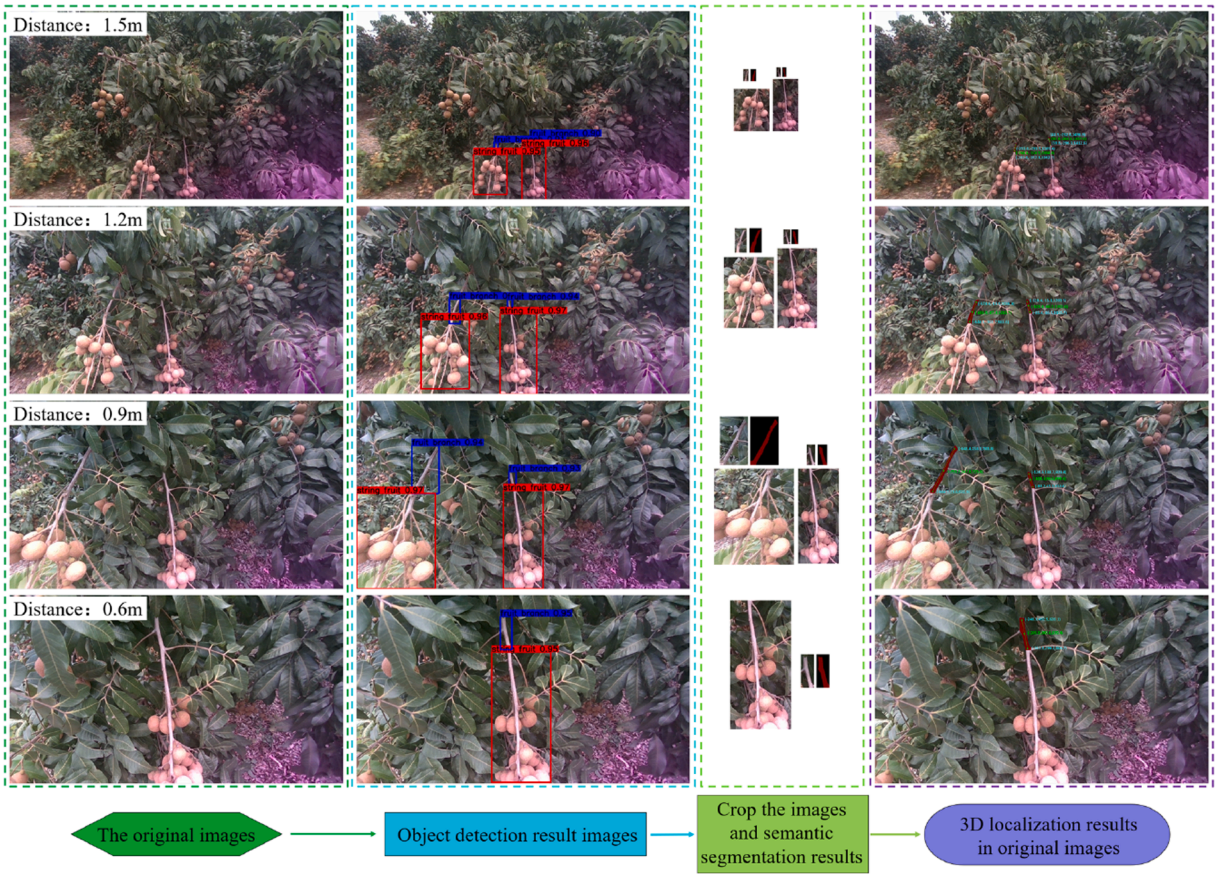


Fig. 22. Experimental results of the overall algorithm in different distance scenes.

Table 4
Statistical results of 3D localization and error in different distance scenes.

Distance		1.5 m	1.2 m	0.9 m	0.6 m
The upper (mm)	Measured value	(−251.3,−291.1,1395.8) (43.2,−231.2,1503.6)	(−575.7,−16.0,1042.0) (−129.5,−15.9,1198.7)	(−646.9,259.6,766.6) (−123.8,139.5,927.4)	(−245.9,493.1,627.4)
	Predicted value	(−253.8,−293.5,1389.4) (41.9,−232.8,1496.8)	(−573.6,−14.9,1046.3) (−128.4,−15.1,1203.5)	(−648.4,258.5,763.8) (−124.6,138.7,923.8)	(−246.3,492.5,626.1)
	Error	(2.5,2.4,6.4) (1.3,1.6,6.8	(2.1,1.1,4.3) (1.1,0.8,5.2)	(1.5,1.1,2.8) (0.8,0.8,3.6)	(0.4,0.6,1.3)
The lower (mm)	Measured value	(−284.4,−354.7,1336.4) (61.1,−294.0,1419.1)	(−650.5,−163.2,980.0) (−86.2,−87.6,1181.1)	(−862.8,24.2,695.4) (−82.6,46.0,856.9)	(−163.6,294.6,605.5)
	Predicted value	(−281.6,−352.1,1342.7) (59.6,−296.3,1412.6)	(−648.1,−161.7,983.6) (−85.4,−86.5,1185.7)	(−864.7,23.6,692.9) (−83.2,45.3,853.6)	(−163.8,294.2,604.7)
	Error	(2.8,2.6,6.3) (1.5,2.3,6.5)	(2.4,1.5,3.6) (0.8,1.1,4.6)	(1.9,0.6,2.5) (0.6,0.7,3.3)	(0.2,0.4,0.8)
The center (mm)	Measured value	(−271.0,−327.4,1358.4) (52.2,−262.2,1464.1)	(−606.2,−88.6,1011.5) (−104.7,−47.1,1187.9)	(−751.6,148.9,728.1) (−100.6,91.4,901.9)	(−205.6,390.1,620.5)
	Predicted value	(−268.4,−324.9,1364.8) (50.8,−264.1,1457.4)	(−603.9,−87.3,1015.7) (−103.8,−46.2,1192.6)	(−753.2,147.8,725.4) (−101.3,90.6,898.4)	(−205.9,389.6,619.4)
	Error	(2.6,2.5,6.4) (1.4,1.9,6.7)	(2.3,1.3,4.2) (0.9,0.9,4.7)	(1.6,1.1,2.7) (0.7,0.8,3.5)	(0.3,0.5,1.1)

To locate picking points from these images, a scheme of quickly locating picking points based on DL is proposed.

First, the collected longan images are processed by image preprocessing methods such as image cropping and image normalization, and the initial and semantic segmentation longan datasets are constructed. Second, to reduce the computations, memory occupation and detection time of the airborne microprocessor, the performance of five object detection models is compared. It is determined that the improved YOLOv5s model can quickly and accurately detect the string fruit and the main fruit branch. Additionally, to improve the efficiency and accuracy of semantic segmentation, the main fruit branch is extracted as the input image for semantic segmentation. Then, the lightweight MobileNetv2 is used as the backbone network to improve the DeepLabv3 + model, which not only reduces the number of calculations and identification time of the model but also improves the semantic segmentation precision. Finally, combining the semantic segmentation results with the depth image, a scheme of estimating the position and posture of the main fruit branches in 3D space and accurately locating the picking points is developed, and the positioning results of the picking points are analysed. In summary, this paper fully exploits the advantages of the combination of a CNN and RGB-D camera. It is more suitable for performing object detection, semantic segmentation and 3D positioning

tasks on the onboard microprocessor of longan harvesting UAVs and improves the speed and accuracy of picking point positioning of longan harvesting UAVs based on visual perception. The research in this paper can apply UAV migration to harvesting string fruits such as grapes, Cerasus pseudocerasus and Lycopersicon esculentum. Additionally, the object positioning method based on a convolutional neural network and RGB-D camera proposed in this research will further provide guidance for developing intelligent harvesting robots in the agricultural field and promote developing intelligent agriculture and unmanned farms.

Considering the complexity of dynamic changes in longan orchard scenes and the influence of abnormal weather conditions, there are still some limitations in this work, and the ability of the object detection, semantic segmentation and 3D positioning schemes to resist environmental interference needs to be further improved. In future work, we will continue to optimize the details of the solution and promote the intelligence of longan harvesting drones.

CRediT authorship contribution statement

Denghui Li: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Data curation, Visualization, Validation. **Xiaoxuan Sun:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Data curation, Visualization, Validation. **Shengping Lv:** Methodology, Software, Validation. **Hamza Elkhouchlaa:** Validation, Writing – review & editing. **Yuhang Jia:** Validation, Data curation. **Zhongwei Yao:** Validation, Data curation. **Peiyi Lin:** Validation, Data curation. **Haobo Zhou:** Validation, Data curation. **Zhengqi Zhou:** Validation, Data curation. **Jiaxing Shen:** Conceptualization, Supervision, Validation. **Jun Li:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

Funding: This work is supported by the earmarked fund for the Laboratory of Lingnan Modern Agriculture Project (NZ2021009), the China Agriculture Research System of MOF and MARA (No. CARS-32-13), the Special Project of Rural Vitalization Strategy of Guangdong Academy of Agricultural Sciences (No. TS-1-4), and the Guangdong Provincial Modern Agricultural Industry Technology System (No. 2021KJ123).

References

Anagnostis, A., Tagarakis, A.C., Asiminari, G., Papageorgiou, E., Kateris, D., Moshou, D., Bochtis, D., 2021. A deep learning approach for anthracnose infected trees classification in walnut orchards. Comput. Electron. Agric. 182, 105998.

Baheti, B., Innani, S., Gajre, S., Talbar, S., 2020. Semantic scene segmentation in unstructured environment with modified DeepLabV3+. Pattern Recogn. Lett. 138 (4), 223–229. https://doi.org/10.1016/j.patrec.2020.07.029.

Bargoti, S., Underwood, J., 2016. Deep Fruit Detection in Orchards. IEEE. 3626–3633. https://doi.org/10.1109/ICRA.2017.7989417.

Bochkovskiy, A., Wang, C.Y., Liao, H., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. https://doi.org/10.48550/arXiv.2004.10934.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *Ieee Trans. Pattern Analy. Machine Intell. 40*(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. https://doi.org/10.48550/arXiv.1706.05587.

da Silva, C.B., Bianchini, V.d.J.M., Medeiros, A.D.d., Moraes, M.H.D.d., Marassi, A.G., Tannús, A., 2021. A novel approach for Jatropha curcas seed health analysis based on multispectral and resonance imaging techniques. Ind. Crops Prod. 161, 113186.

de Medeiros, A.D., Bernardes, R.C., da Silva, L.J., de Freitas, B.A.L., Dias, D.C.F.D.S., da Silva, C.B., 2021. Deep learning-based approach using X-ray images for classifying Crambe abyssinica seed quality. Ind. Crops Prod. 164, 113378.

Dyrmann, M., Jørgensen, R.N., Midtiby, H.S., 2017. RoboWeedSupport - Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. Adv. Anim. Biosci. 8 (2), 842–847.

Feng, A., Zhou, J., Vories, E., Sudduth, K.A., 2020a. Evaluation of cotton emergence using UAV-based imagery and deep learning. Comput. Electron. Agric. 177, 105711.

Feng, A.J., Zhou, J.F., Vories, E.D., Sudduth, K.A., Zhang, M.N., 2020b. Yield estimation in cotton using UAV-based multi-sensor imagery. Biosyst. Eng. 193, 101–114. https://doi.org/10.1016/j.biosystemseng.2020.02.014.

Flores, P., Zhang, Z., Igathinathane, C., Jithin, M., Naik, D., Stenger, J., Ransom, J., Kiran, R., 2021. Distinguishing seedling volunteer corn from soybean through greenhouse color, color-infrared, and fused images using machine and deep learning. Ind. Crops Prod. 161, 113223.

Fu, L.S., Majeed, Y., Zhang, X., Karkee, M., Zhang, Q., 2020. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. Biosyst. Eng. 197, 245–256. https://doi.org/10.1016/j.biosystemseng.2020.07.007.

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., Zhang, Q., 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 176, 105634.

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. An explainable deep machine vision framework for plant stress phenotyping. PNAS 115 (18), 4613–4618. https://doi.org/10.1073/pnas.1716999115.

Girshick, R., 2015. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2016. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 38 (1), 142–158. https://doi.org/10.1109/TPAMI.2015.2437384.

He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2015. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916. https://doi.org/10.1109/tpami.2015.2389824.

Jaisin, C., Pathaveerat, S., Terdwongworakul, A., 2013. Determining the size and location of longans in bunches by image processing technique. Maejo Int. J. Sci. Technol. 7 (3), 444–455. 10.14456/mijst.2013.37.

Kamilaris, A., Prenafeta-Boldu, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kang, H., Chen, C., 2020. Fruit detection, segmentation and 3D visualisation of environments in apple orchards. Comput. Electron. Agric. 171, 105302.

Koirala, A., Walsh, K.B., Wang, Z.L., McCarthy, C., 2019. Deep learning - Method overview and review of use for fruit detection and yield estimation. Comput. Electron. Agric. 162, 219–234. https://doi.org/10.1016/j.compag.2019.04.017.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Li, D., Sun, X., Elkhouchlaa, H., Jia, Y., Yao, Z., Lin, P., Li, J., Lu, H., 2021. Fast detection and location of longan fruits using UAV images. Comput. Electron. Agric. 190, 106465.

Li, J.H., Tang, Y.C., Zou, X.J., Lin, G.C., Wang, H.J., 2020. Detection of Fruit-Bearing Branches and Localization of Litchi Clusters for Vision-Based Harvesting Robots. IEEE Access 8, 117746–117758. https://doi.org/10.1109/ACCESS.2020.3005386.

Liang, C., Xiong, J., Zheng, Z., Zhong, Z., Li, Z., Chen, S., Yang, Z., 2020. A visual detection method for nighttime litchi fruits and fruiting stems. Comput. Electron. Agric. 169, 105192.

Lin, P., Kan, K.-W., Chen, J.-H., Lin, Y.-K., Lin, Y.-H., Lin, Y.-H., Hu, W.-C., Chiang, C.-F., Kuan, C.-M., Fioravanti, A., 2020. Investigation of the Synergistic Effect of Brown Sugar, Longan, Ginger, and Jujube (Brown Sugar Longan Ginger Tea) on Antioxidation and Anti-Inflammation in In Vitro Models. Evidence-Based Complement. Alternative Med. 2020, 1–6.

Lin, T.Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017, 936–944. https://doi.org/10.1109/CVPR.2017.106.

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path Aggregation Network for Instance Segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018, 8759–8768. https://doi.org/10.1109/CVPR.2018.00913.

Liu, Z.H., Wu, J.Z., Fu, L.S., Majeed, Y., Feng, Y.L., Li, R., Cui, Y.J., 2020. Improved Kiwifruit Detection Using Pre-Trained VGG16 With RGB and NIR Information Fusion. IEEE Access 8, 2327–2336. https://doi.org/10.1109/ACCESS.2019.2962513.

Ma, L., Liu, Y., Zhang, X.L., Ye, Y.X., Yin, G.F., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. PNAS 115 (25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115.

Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: A review. ISPRS J. Photogramm. Remote Sens. 158, 279–317. https://doi.org/10.1016/j.isprsjprs.2019.09.006.

Pham, V.T., Herrero, M., Hormaza, J.I., 2015. Phenological growth stages of longan (Dimocarpus longan) according to the BBCH scale. Sci. Hortic. 189, 201–207. https://doi.org/10.1016/j.scienta.2015.03.036.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, 779–788. https://doi.org/10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement. *arXiv e-prints*. https://doi.org/10.48550/arXiv.1804.02767.

Ren, S.Q., He, K.M., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Ronneberger, O., Fischer, P., Brox, T., 2015. In: U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer International Publishing, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Singh, P., Verma, A., Alex, J.S.R., 2021. Disease and pest infection detection in coconut tree through deep learning techniques. Comput. Electron. Agric. 182, 105986.

Sumesh, K.C., Ninsawat, S., Som-ard, J., 2021. Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. Comput. Electron. Agric. 180, 105903.

Tetila, E.C., Machado, B.B., Menezes, G.K., Oliveira, A.D., Alvarez, M., Amorim, W.P., Belete, N.A.D., da Silva, G.G., Pistori, H., 2020. Automatic Recognition of Soybean Leaf Diseases Using UAV Images and Deep Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 17 (5), 903–907. https://doi.org/10.1109/LGRS.2019.2932385.

Vanegas, F., Bratanov, D., Powell, K., Weiss, J., Gonzalez, F., 2018. A Novel Methodology for Improving Plant Pest Surveillance in Vineyards and Crops Using UAV-Based Hyperspectral and Spatial Data. Sensors 18 (1). https://doi.org/10.3390/s18010260.

Xiong, J.T., He, Z.L., Lin, R., Liu, Z., Bu, R.B., Yang, Z.G., Peng, H.X., Zou, X.J., 2018. Visual positioning technology of picking robots for dynamic litchi clusters with disturbance. Comput. Electron. Agric. 151, 226–237. https://doi.org/10.1016/j.compag.2018.06.007.

Xiong, J.T., Liu, Z., Chen, S.M., Liu, B.L., Zheng, Z.H., Zhong, Z., Yang, Z.G., Peng, H.X., 2020. Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method. Biosyst. Eng. 194, 261–272. https://doi.org/10.1016/j.biosystemseng.2020.04.006.

Zhang, J., Karkee, M., Zhang, Q., Zhang, X., Yaqoob, M., Fu, L., Wang, S., 2020. Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting. Comput. Electron. Agric. 173 https://doi.org/10.1016/j.compag.2020.105384.

Zhong, Z., Xiong, J., Zheng, Z., Liu, B., Liao, S., Huo, Z., Yang, Z., 2021. A method for litchi picking points calculation in natural environment based on main fruit bearing branch detection. Comput. Electron. Agric. 189 https://doi.org/10.1016/j.compag.2021.106398.

Zhou, J., Zhou, J., Ye, H., Ali, M.L., Nguyen, H.T., Chen, P., 2020. Classification of soybean leaf wilting due to drought stress using UAV-based imagery. Comput. Electron. Agric. 175 https://doi.org/10.1016/j.compag.2020.105576.

Zhuang, J.J., Hou, C.J., Tang, Y., He, Y., Guo, Q.W., Zhong, Z.Y., Luo, S.M., 2019. Computer vision-based localisation of picking points for automatic litchi harvesting applications towards natural scenarios. Biosyst. Eng. 187, 1–20. https://doi.org/10.1016/j.biosystemseng.2019.08.016.