IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 22, 2020, accepted October 2, 2020, date of publication October 6, 2020, date of current version November 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029215

Fast and Accurate Detection of Banana Fruits in Complex Background Orchards

LANHUI FU[1], JIELI DUAN[1], XIANGJUN ZOU[1], JIAQUAN LIN[1], LEI ZHAO[1], JINHUI LI[1], AND ZHOU YANG[1,2]

[1]College of Engineering, South China Agricultural University, Guangzhou 510642, China
[2]Guangdong Provincial Key Laboratory of Conservation and Precision Utilization of Characteristic Agricultural Resources in Mountainous Areas, Jiaying University, Meizhou 514015, China

Corresponding authors: Jieli Duan (duanjieli@scau.edu.cn) and Zhou Yang (yangzhou@scau.edu.cn)

This work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2019B020223002, in part by the National Natural Science Foundation of Guangdong Province under Grant 2020A1515010793, and in part by the Major Scientific Research Projects of Guangdong Province under Grant 2020KZDZX1037.

ABSTRACT The detection of banana fruits is an important part of intelligent management in the banana plantation. To detect the banana fruit quickly and accurately in the complex orchard environment, this article proposes a method based on the latest deep learning algorithm to detect the banana fruit. Using a monocular camera, we applied the YOLOv4 neural network algorithm to extract the deep features of banana fruits, realizing accurate detection of different banana sizes. The detection algorithm achieved a 99.29% detection rate, the average execution time was 0.171s, the shortest execution time was 0.135s, and the AP was 0.9995. Moreover, the detection results were discussed with the YOLOv3 algorithm and the machine learning algorithm. Compared with the machine learning algorithm, deep learning algorithm was superior to both detection accuracy and detection time. YOLOv4 had higher detection confidence and higher detection rate than YOLOv3. The results show that the proposed method could realize the fast detection of different varieties and different maturity in banana plantations, under different illumination and occlusion conditions, and provide information for banana picking, maturity and yield estimation.

INDEX TERMS Banana detection, orchard environment, deep learning, green fruit, YOLOv4.

I. INTRODUCTION

Robots have attracted wide attention in the field of agriculture, with the shortage of farm labor and the rapid development of artificial intelligence. Agricultural robots automate tedious farm work and enable farmers to better focus on farm management. The harvesting robot is one of the most popular agricultural robots. In recent years, harvest robots have significantly improved on speed and accuracy, and people are increasingly interested in agricultural robots to harvest fruits and vegetables. Visual system is the key to realize automatic harvest, and accurate detection is the premise of follow-up operation and picking in fruit and vegetable harvest. However, it is a great challenge to achieve a robust and efficient fruit detection algorithm, due to the similarity or occlusion of fruits and branches and other background problems as well as the uncertainty of the orchard environment.

Banana is the world's most popular fruit and an important source of staple food. Due to the irregular shape and green color of banana, accurate detection becomes the primary task of banana harvesting robot in the natural environment. This article presents a detection method of the banana in the plantation based on deep learning algorithm. In this method, we used a regular RGB color camera to obtain banana images in the plantation, and banana fruits were detected under different illumination and occlusion conditions. The main research contents include: (1) Based on the latest detection algorithm YOLOv4 [1], fast and accurate detection of banana fruits can be realized under various environmental conditions; (2) The results in this article were discussed with the banana detection results based on YOLOv3 [2] algorithm and machine learning algorithm, to verify the applicability and high efficiency of the proposed method in banana detection.

The rest of this article is structured as follows. The second part reviews the related work. The third part introduces the structure and implementation of the banana detection algorithm in the plantation. Part four and part five introduce the

The associate editor coordinating the review of this manuscript and approving it for publication was Hongjun Su.

experimental results and comparative discussion. In the sixth part, it includes the summary and the plan of future work.

II. RELATED WORK

In this section, we first review the development of convolutional neural networks in deep learning algorithms and then discuss the research on fruit and vegetable detection. Besides, the research progress of our topic is introduced.

A. DEVELOPMENT OF CONVOLUTIONAL NEURAL NETWORKS

In the field of deep learning, the convolutional neural network algorithm can be divided into three categories according to the research purpose: classification networks, detection networks, and segmentation networks.

Among classification networks, LeNet [3] is one of the earliest convolutional neural networks. In 2012, Alex *et al.* [4] deepened the network structure based on LeNet and learned higher-dimensional image features. In 2014, Karen and Andrew [5] proposed VGGNet and successfully constructed a convolutional neural network with 16-19 layers of depth, proving that increasing the depth of the network could affect the ultimate performance of the network. In 2015, He *et al.* [6] trained the 152-layer deep neural network with Residual Unit, achieving a 3.57% top5 error rate, whereas the number of parameters was lower than VGGNet. From 2014 to 2016, Google proposed convolutional networks Inception v1-v4 [7]–[10]. Compared to VGGNet, Inception-v1 changed the full connection and convolutional layer to a sparse connection. Inception-v2 proposed Batch Normalization. Inception-v3 increased network depth and nonlinearity, and the network input was changed from 224 × 224 to 299 × 299. Inception-v4 combined the Inception and the ResNet. In 2016, Chollet [11] presented Xception, introduced depthwise separable convolution based on the Inception v3, the model was improved without increasing the network complexity.

The detection networks are divided into two categories, one is based on the candidate regions (two-stage detector), and the other is based on the regression method (one-stage detector). **Representative networks based on candidate regions are the R-CNN series.** In 2013, R-CNN was proposed by Girshick *et al.* [12], and feature vectors were extracted from each region proposals using CNN, then linear SVM was used for classification. In April 2015, Girshick [13] presented Fast R-CNN, adopted the selective search method to achieve a higher object detection accuracy of the model. In June 2015, Ren *et al.* [14] used RPN (Region Proposal Network) instead of selective search to produce a proposal window, which greatly improved efficiency and this method is called Faster R-CNN. In 2016, Lin *et al.* [15] further improved the Faster R-CNN, proposed FPN, amplified coarse outputs, and fine-tuned the outputs with a convolution feature map to get better results. **Regression-based representative networks are the YOLO series and SSD.** In 2015, Redmon *et al.* [16] proposed YOLO, divided the image into S × S grid, predicted the bounding boxes, the confidence, and the probability of all categories of the objects in all cells in one-shot. In 2016, Liu *et al.* [17] presented SSD (Single Shot MultiBox Detector), aiming at the weaknesses and advantages of YOLOv1 and Faster R-CNN. In the same year, Redmon and Farhadi [18] used anchor based on YOLOv1 and SSD, thus putting forward YOLOv2, which improved the performance. In 2018, YOLOv3 was released [2], using multi-scale feature detection and logistic instead of softmax for classification, which improved the accuracy and ensured speed. In 2020, Bochkovskiy *et al.* [1] presented the latest version YOLOv4, summarized almost all detection tricks, and developed an object detection model with faster speed and better accuracy, which is better than previous versions in small object detection and occlusion object detection. Therefore, it was used in this study.

The segmentation networks have the semantic segmentation networks and the instance segmentation networks. In 2014, Long *et al.* [19] proposed FCN to classify images at the pixel level. In 2015, Badrinarayanan *et al.* [20] proposed Segnet, using deconvolution and upper pooling. In 2014, Chen *et al.* [21] presented DeepLab-v1 based on VGG16. DeepLab-v2 [22] was proposed in 2016 that the base layer was transformed from VGG16 to ResNet to achieve a better segmentation effect with multiple scales. In 2017, a more generic framework DeepLab-v3 [23] was released, replicating the last block in ResNet, using the BN layer in ASPP. In 2018, DeepLab-v3 + [24] appeared, based on the decode module and modified Xception as the backbone. Liu *et al.* [25] proposed Auto-DeepLab in 2019, which could search effectively on a two-level hierarchical architecture. Mask RCNN [26], an instance segmentation network, was proposed by He *et al.* in 2017, taking the Faster R-CNN as the prototype and the ResNet-FPN architecture for feature extraction, it can be used for human attitude estimation and other tasks.

B. RESEARCH ON FRUIT AND VEGETABLE DETECTION

Fruit and vegetable detection is one kind of object detection. Object detection based on vision technology has a myriad of applications in various engineering fields [27]–[29]. Traditional object detection algorithms are based on hand-designed features (such as color, shape, texture, strength, or fusion features) and appropriate classifiers (Support vector machine, Adboost, etc.) to locate the region of interest in the image. These methods often lack universality and robustness. With the development of deep learning technology, the application of deep convolutional neural networks for fruit and vegetable detection has been the focus of research in recent years. Deep learning can extract deep features and have stronger learning ability. These algorithms have been shown to detect fruits and vegetables in uncontrolled environments.

In the study of fruit detection, apple fruit detection and branch segmentation are the focus of researchers [30]–[33]; The establishment of a dedicated neural network for mango detection continues to emerge [34]–[37]; Various neural networks in litchi [38], [39], grape [40], [41], strawberry [42], [43] have achieved good results in their application. The detection of pomelo [44], kiwi fruit [45], waxberry [46], guava [47], and other fruits have been gradually concerned; With the development of deep learning, fruit flower detection, which is difficult to the traditional algorithm, has been emerging [48]–[51]. In the detection of vegetables, the improvement in the bounding box and the detection rate is the research focus of the tomato detection network [52]–[54]; Based on deep neural network, excellent results have been achieved in cucumber fruit length estimation [55], sweet pepper detection [56], date fruit variety and maturity judgment [57] and other aspects.

In recent years, deep convolutional neural network has been applied in the banana plantation. Based on fast-RCNN, Neupane *et al.* [57] recognized and counted banana plants on the farm by using RGB aerial images collected by UAV. Clark and McKechnie [58] detected banana plantations through aerial images and used U-NET neural network to draw maps, but did not conduct detection and research on banana fruits. Chen *et al.* [59] detected the banana central stocks using Deeplab V3 + network with two binocular cameras, and obtained satisfactory results.

C. RESEARCH PROGRESS OF OUR TOPIC

In our early work [60], we demonstrated that using traditional machine learning algorithm SVM classifier with color and texture features can achieve impressive results in banana detection. However, early work focused on detecting orchard bananas of the same variety for CPU processing. When different varieties appeared, the early-trained banana detection model could be used as the basis for this phase. Meanwhile, GPU processing capability provides support for faster and more efficient detection.

The advantage of our approach is that we use a regular RGB camera instead of the complicated sensors, which greatly reduces the cost of collecting images of banana fruits. In this work, we introduce the latest and most powerful detection algorithm YOLOv4, which is used to identify the key features of the banana image and find the banana fruit, to imitate the human eye for the rapid detection of banana fruits in the plantation.

III. MATERIALS AND METHODS

A. IMAGE ACQUISITION

For developing and testing the proposed algorithm, 388, 178 and 134 valid banana images were acquired at the banana plantation of Guangdong Academy of Agricultural Sciences on August 9, 2018 (sunny), November 19, 2018 (cloudy) and March 16, 2019 (overcast), 464 valid banana images were acquired at Nansha banana plantation in Guangzhou on October 27, 2019 (sunny). A digital color camera (Canon sx610hs) with a resolution of 2048 × 1536 pixels is used. The camera exposure mode was set to auto exposure, the camera height was 150 cm, the shooting distance was about 80 – 120 cm, and the shooting angle was set to horizontal. In addition, 10 photos with an elevation angle of 45°– 60° were taken for comparison. In the 1164 images. The training set, validation set, and test set were 835, 209, and 120 images respectively. We used Python (PyCharm Community Edition 2019.3.1 x64) to implement the algorithm on an Intel(R) Core(TM) i7 – 9750H @2.6 GHz 2.59GHz, 16.0 GB RAM, NVIDIA GeForce RTX 2070 with Max-Q Design laptop. Colabeler, a free and open-source labeling tool, was used to label each image. Once the fruits are labeled, an Extensible Markup Language (XML) file is generated that contains the label data and the coordinates of the bounding box for each fruit in the image.

B. ALGORITHM DESCRIPTION

The latest detection algorithm YOLOv4 is applied in this article. YOLO series is favored by researchers for the flexible structure and rapid detection. With the continuous optimization of the algorithm, YOLOv4 combines a large number of tricks to achieve faster speed and better accuracy. In the following part, the internal structure of the network is introduced in detail, and the applicability of the network in banana detection in the orchard is explained from the principle and structural design.

Fig. 1 is the flow chart of banana detection based on YOLOv4 algorithm. The detection process is as follows:

Step 1: A banana image is fed into the network.

Step 2: The backbone is a CSPDarknet53 module and the Mish activation function is adopted, which extracts the information from the image.

Step 3: The neck part is composed of SPP (Spatial Pyramid Pooling) module and FPN (Feature Pyramid Networks) + PAN (Path Aggregation Network) module, which is to make better use of the characteristic extracted by the backbone.

Step 4: The head is the prediction part, which uses the features extracted earlier and outputs the final detection result. Next, we elaborate on the contents of the specific module.

The backbone is a CSPDarknet53 structure, consisting of 5 CSP (Cross Stage Partial connections) modules (blue block) and 11 CBM (Convolutional+ Batch normalization + Mish) modules (yellow block). The CBM module represents a convolution operation that uses the Batch Normalization and Mish activation functions. The CBM module is an important part of the CSP module. The CSP module will be explained in detail below. Similar to the CBM module, the CBL (Convolutional+ Batch normalization + Leaky Relu) module (green block) represents another type of convolution operation that uses the Batch Normalization and Leaky Relu activation functions.

It is worth mentioning that, in order to get access to a much richer hypothesis space that would benefit from deep representations, researchers need the activation function to generate nonlinear mappings between inputs and outputs. Leaky Relu function is a popular activation function in deep learning, and the average performance of the Mish function is better than that of the Leaky Relu function. The use of

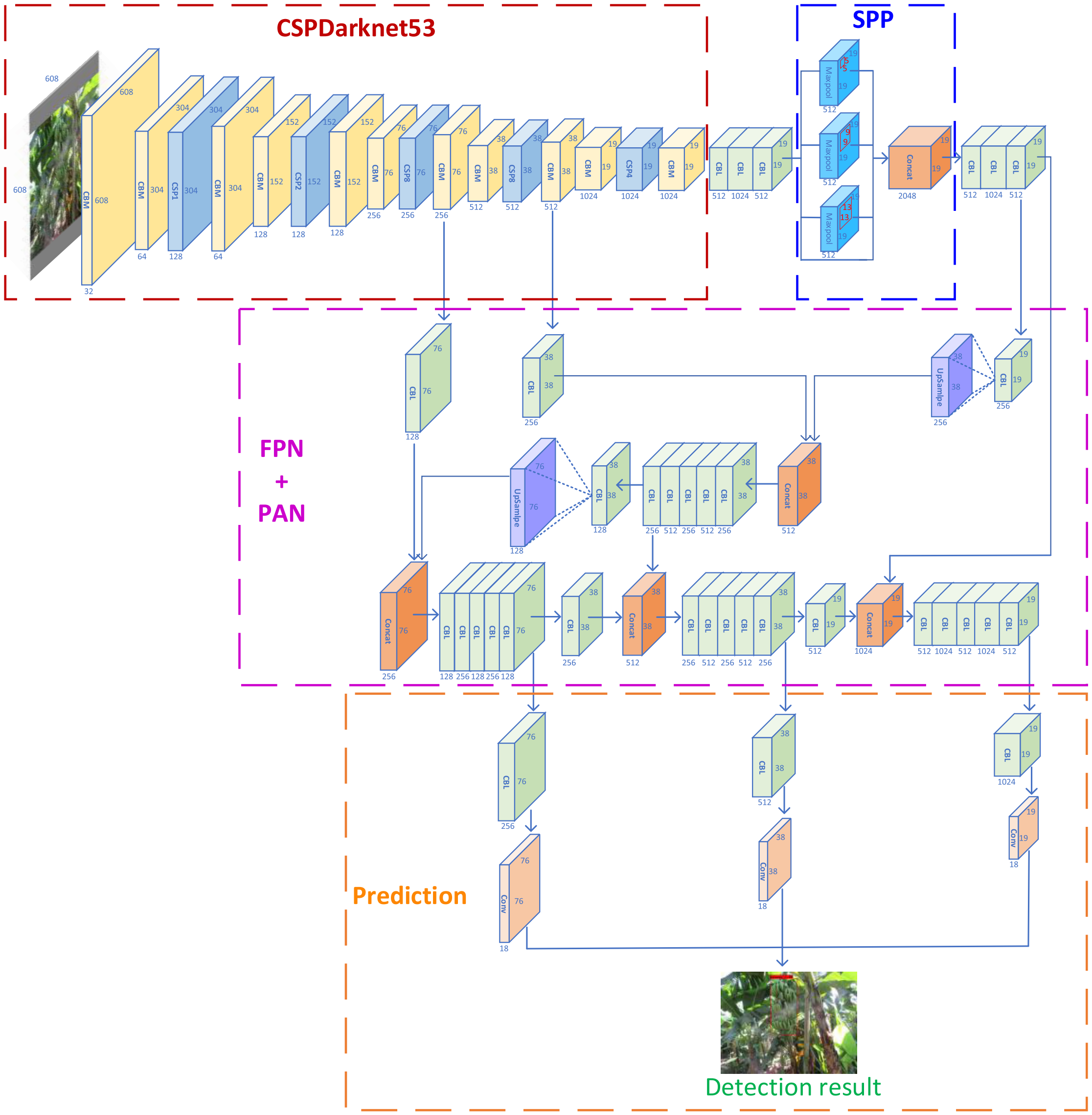


FIGURE 1. Flow chart of banana detection based on YOLOv4.

Mish activation function is one of the innovations of the network, which can improve the detection accuracy. The network adopts the Mish activation function over the backbone, and the Leaky Relu activation function remains throughout the rest of the network. Mish function is,

$$y_{mish} = x \tanh(\ln(1 + \mathrm{e}^{x})). \tag{1}$$

Leaky relu function is,

$$y_{leaky\ relu} = \begin{cases} x, & if\ x \geq 0 \\ \lambda x, & if\ x < 0. \end{cases} \tag{2}$$

The graphs of Mish function and Leaky Relu function are compared as shown in Fig.2.

The use of the CSP module is one of the network innovations. CSPn is used to represent n Res units in the module. The structure is shown in Fig. 3, where the Add operation is the addition of tensors without extending dimensions, and the Concat operation is the addition of tensors and dimensions.

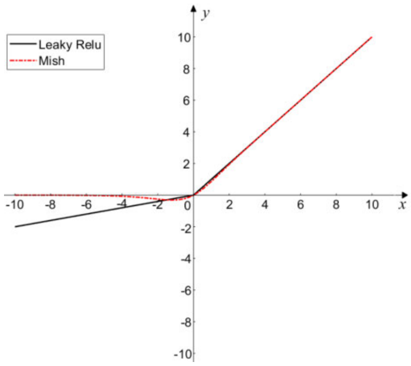


FIGURE 2. The Mish function and the Leaky Relu function.

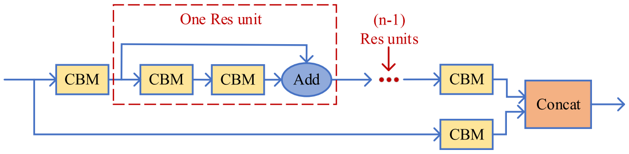


FIGURE 3. CSPn structure.

CSP1 means one Res unit; CSP8, similarly, means 8 Res units. After five CSP modules, the size of the input image is gradually changed from 608 to 19 through down-sampled. From the structure diagram, the CSP module maps the upper features into two parts for different convolution operations and then merges them to reduce memory cost and ensure accuracy. The introduction of the Res Unit makes the network deeper and more features can be extracted. Banana in the plantation is green fruit, which color is very close to that of banana stem, branches, and leaves. The shape of banana fruits is irregular. The background has a great interference on banana detection. Therefore, it is very important to extract the deep features of banana fruits.

The neck structure in the network adopts the SPP module (cyan) and the FPN+PAN module (purple dotted line area). In the SPP module, $1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13$ max-pooling are adopted, the padding is 2, the stride is 1, to ensure the size remains unchanged after pooling. Compared with the traditional max-pooling method, the SPP module can increase the acceptance range of backbone network features and achieve more accuracy improvement with less calculation cost. Based on the use of the FPN module in YOLOv3, YOLOv4 added the PAN module, which is another structural innovation. Following the arrow direction in the figure, it can be seen that FPN amplifies the size of the feature map through up-sampling operation, to fuse tensor and dimension with the feature map after CSP operation in the backbone network, and convey object semantic information. After down-sampled the fused feature map through convolution operation, the PAN structure is fused with the feature map of the corresponding scale in FPN to further extract positioning features. FPN+PAN fuses different trunk layers and detection layers repeatedly and uses multiple scales to extract more profound semantic information and positioning information, to detect more delicate objects of different sizes. As a result, the detection of small objects has been greatly improved. In banana detection, the fruit sizes of different varieties vary greatly. When bananas of different sizes appear in the same image, the generalization ability of the detection algorithm is very important, which will be explained in the discussion section.

The head structure in the network is the prediction part (orange dotted line area), through the CBL module and convolution operation, the three-layer scale feature maps (76×76, 38×38, 19×19) obtained from the upper network are output. Each scale predicts three anchor boxes, and there are 6 values per anchor (4 box coordinates + 1 object confidence + 1 class confidences). Therefore, each layer has 18 outputs. The bounding box and its confidence of the detected banana can be obtained according to the output information. Then, the bounding box whose confidence is lower than the threshold is deleted, and the best candidate box would be selected according to the *DIOU_nms* algorithm. The adoption of *DIOU* (Distance-IOU) is the innovation of network structure. The calculation formula of *DIOU* is,

$$DIOU = IOU - \frac{\alpha^2}{\beta^2}. \tag{3}$$

where, α represents the distance between the center points of the two boxes, and β represents the diagonal distance between the minimum closure areas of the two boxes. DIOU_nms believes that boxes with far central points may be on different objects and should not be deleted, which is the biggest difference between DIOU_nms and traditional NMS. After filtering by DIOU_nms, the detection result is output, and the detection task is finished.

IV. RESULT

This section explains the results of banana detection in the training stage and detection stage. The evaluation indexes, training parameters and detection effects in different scenes are described.

A. EVALUATION OF TRAINING MODELS

In the training stage, to evaluate the generalization ability and gradually optimize the model, precision, recall and the F_1 score were used as evaluation indexes:

$$Precision = \frac{Truepositive}{Truepositive + Falsepositive} \times 100\% \tag{4}$$

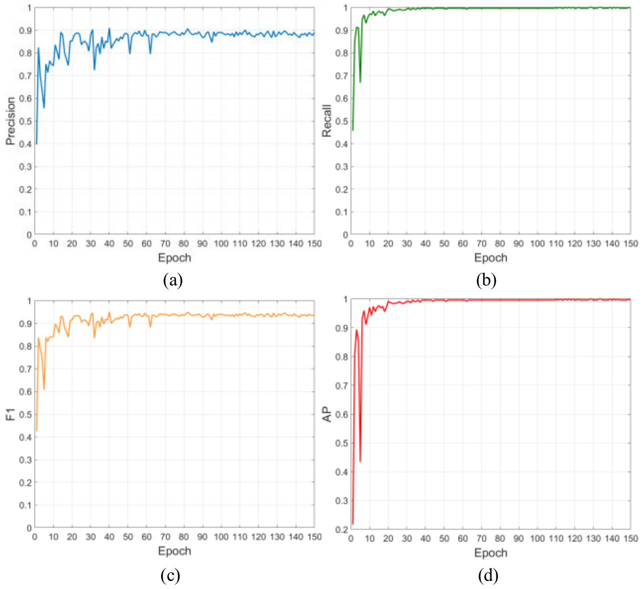
$$Recall = \frac{Truepositive}{Truepositive + Falsenegative} \times 100\% \tag{5}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

In the training, the batch size was set to 4, that is, 4 images were taken each iteration, and a total of 835 images were trained. Therefore, one epoch required 209 iterations. The weight results of each epoch were verified in the validation set. Based on a threshold, a group of precision and recall of the model could be obtained. When different thresholds are set for the model, multiple groups of precision and recall would be obtained, thus a P-R curve could be drawn, the area of the curve is the *AP* (Average Precision). Three groups of training were set, with the maximum epoch of 100, 150, and 300 respectively. The weight corresponding to the maximum *AP* in each training was selected, and the precision and recall were output when the threshold was 0.5, to compare the performance of the three trainings, as shown in Table 1. As the epoch and the number of iterations increased, the *AP* became higher and took more time. At the group of 300 epochs, the highest *AP* reached 0.9996, but it took 27 hours. By contrast, the group of 150 epochs is enough to get a high *AP* value of 0.9995, which takes 12.5 hours. Therefore, we further analyzed the evaluation indexes in the training process of 150 epochs.

TABLE 1. Performance of the three trainings.

No.	I	II	III
Epoch	100	150	300
Image size	608×608	608×608	608×608
Time(h)	8	12.5	27
Precision	0.8791	0.8796	0.8893
Recall	0.9959	1.0000	1.0000
F_1	0.9339	0.9359	0.9414
AP	0.9957	0.9995	0.9996



(a)

(b)

(c)

(d)

FIGURE 4. Evaluation index curve in 150 epochs training: (a) Precision, (b) Recall, (c) F1, (d) AP.

In the second training, the numerical curves of precision, recall, AP, and F_1 are shown in Fig. 4. It can be seen that the recall curve and the AP curve can converge rapidly and be close to 1. The precision curve and F_1 curve are more stable after the 100th epoch. Therefore, the optimal weight model in the second training is selected as the banana detection model based on the YOLOv4 algorithm.

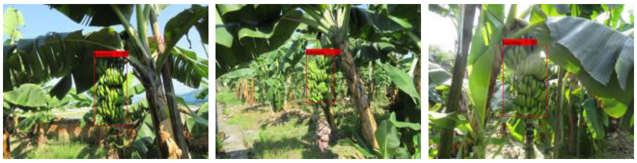


FIGURE 5. Detection results under the sunny front light condition.

B. DETECTION RESULTS

The trained banana detection model was test in different illumination environments. Choose three examples for illustration in each environment. Fig. 5 shows the detection results under the sunny front light condition. Bananas could be detected accurately no matter whether the banana is fully or partially by the light and whether the light is strong or weak. Fig. 6 shows the detection results under the sunny backlight condition. Both one hand of banana and two hands of bananas were accurately detected. Fig. 7 displays the detection results under cloudy conditions. Due to the large size of bananas and the distance between banana plants, it is very rare for more than three hands of bananas to appear in the same image. It's easy to see that each banana in the images was accurately detected under different illumination conditions, which was different from the detection result in [60]. This is because the machine learning algorithm is easily affected by the illumination, while the deep learning algorithm has stronger robustness to the environmental conditions.

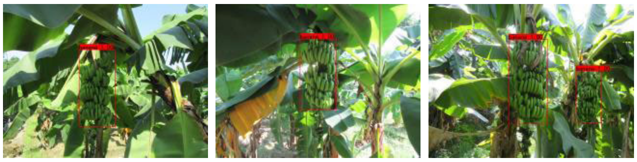


FIGURE 6. Detection results under the sunny backlight condition.

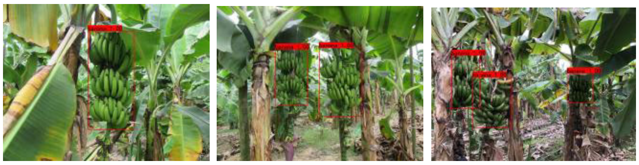
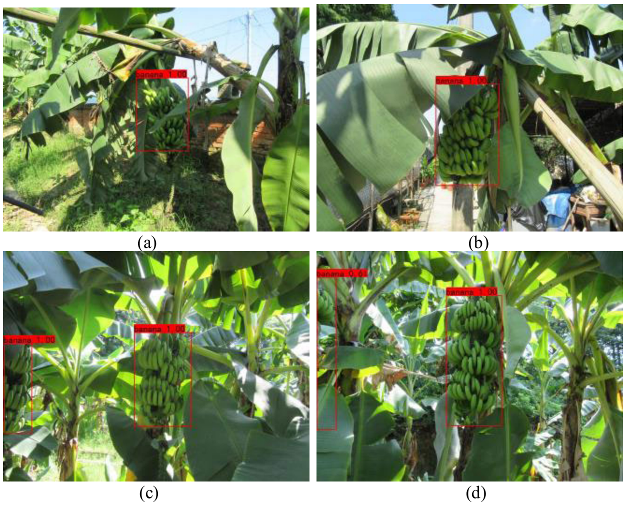


FIGURE 7. Detection results under the cloudy condition.



(a)

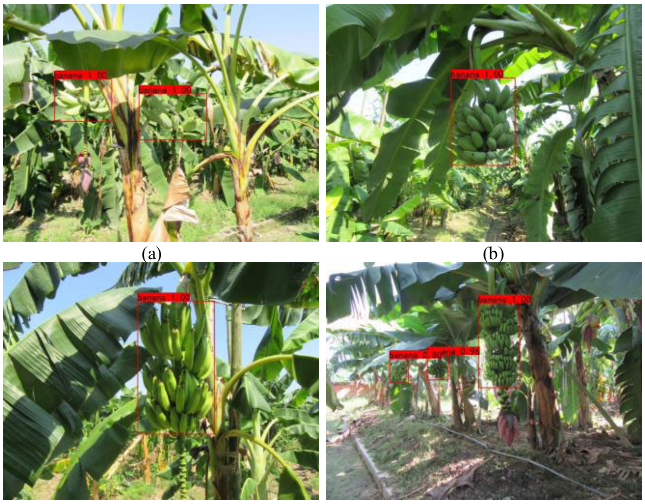
(b)

(c)

(d)

FIGURE 8. Detection results of banana under different occlusion degrees: (a) small region occlusion; (b) occlusion area increased; (c) half of the left banana information was lost; (d) more than half of the left banana information was lost.

Due to the large size of banana branches and leaves, in previous studies, banana detection results were different under different occlusion degrees. At the same time, due to different capture angles, incomplete bananas are easy to appear in one image, which is also classified as an occlusion environment. Therefore, we test on the trained detection model in various occlusion degrees, as shown in Fig. 8: (a) is small region occlusion, that did not affect the detect result; (b) shows the accurate detection when occlusion area increased; half of the left banana in (c) was not captured by the camera, but the model still detected the bananas with a confidence of 1; though the information of the left banana in (d) was almost completely lost, the banana was detected with a confidence of 0.61, that is because the information of the banana was too little. Occlusion in (c) and (d) often occurs in continuous detection. Accurate detection of all bananas in consecutive frames is of great significance for solving the problem of repeated detection.



(a) (b)

(c) (d)

FIGURE 9. Detection results of different varieties of bananas.

There are many varieties of bananas, and new varieties have emerged in recent years. We hope that the banana detection model can realize robust detection for different banana varieties, especially for different varieties and different sizes of bananas in the same scene. Fig. 9 shows the detection results of different varieties of bananas: (a) is the detection result of MUSA AA banana. Although the bananas had poor growth, very few fingers and the light was strong, two hands of bananas were detected; (b) is the detection result of MUSA ABBB banana, which has short and dense finger and no obvious separation in the hand; (c) is the detect result of MUSA ABB banana; (d) is the detection result of one MUSA AAA Cavendish and two MUSA ABBB bananas. Due to the small size of the two hands and their distance from the capture point, the detection confidence is 0.80 and 0.96, respectively.

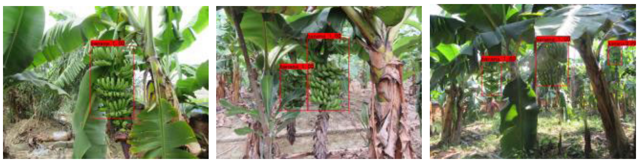


FIGURE 10. Detection results of bananas at different maturity.

Finally, banana fruits at different growing stages were detected, as shown in Fig. 10. Single or multiple hands of immature bananas were correctly detected, including when different maturity bananas were in the same image. It is worth noting that banana confidence is very high, a lot of confidence is 1, which will be further analyzed in comparison with other models in the discussion.

V. DISCUSSION

To verify the performance of the banana detection model in the plantation, other algorithms are compared in this section. Meanwhile, banana detection results based on traditional machine learning algorithm and deep learning algorithm are compared and analyzed.

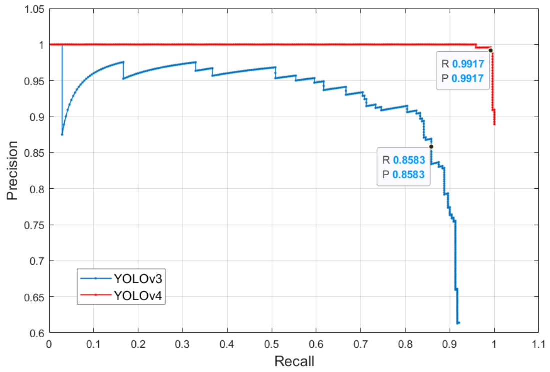


FIGURE 11. P-R curve for different detection methods.

A. COMPARISON OF YOLOv4 AND YOLOv3 IN BANANA DETECTION

We trained and detected the data set in this article in the YOLOv3 neural network. The epoch was set as 300, and the optimal training model was selected for validation, with an *AP* of 0.8697. Fig. 11 shows the P-R curve of the two methods on the validation set. The P-R curve based on the YOLOv3 algorithm is surrounded by the P-R curve of YOLOv4. The break-even point of YOLOv3 is (0.8583,0.8583), and the break-even point of YOLOv4 is (0.9917,0.9917). As mentioned above, the neck part of YOLOv3 uses FPN structure. Different from it, YOLOv4 uses FPN+PAN structure to repeatedly extract the features of the trunk layer and detection layer through multi-scales, which is of great significance for improving network detection of small objects. Since multiple varieties of bananas in the data set were considered, and the distance between banana plants is relatively large, the size of bananas of different varieties or different distances varies greatly in the same image. Therefore, when the banana fruit is very small, YOLOv3 could not detect completely, resulting in a low AP value.

From the detection results, it can be intuitively found that the two detection methods are different in small object detection. Comparison of the detection results of two hands of bananas is shown in Fig. 12, (a) (c) is the detection result of YOLOv3, (b) (d) is the detection result of YOLOv4. YOLOv4 can detect the small banana which was occluded by other banana or by branches and leaves, whereas YOLOv3

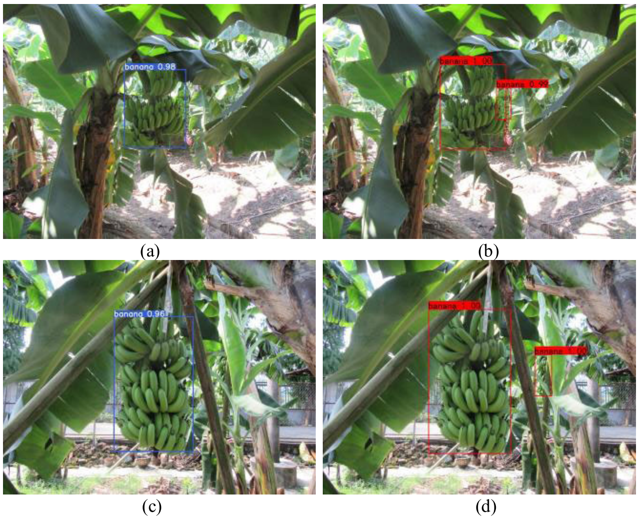


FIGURE 12. Detection results of the two hands of bananas: (a) the detection result of occlusion between banana fruits based on YOLOv3; (b) the detection result of occlusion between banana fruits based on YOLOv4; (c) The detection result of the banana being occluded by leaves based on YOLOv3; (d) The detection result of the banana being occluded by leaves based on YOLOv4.

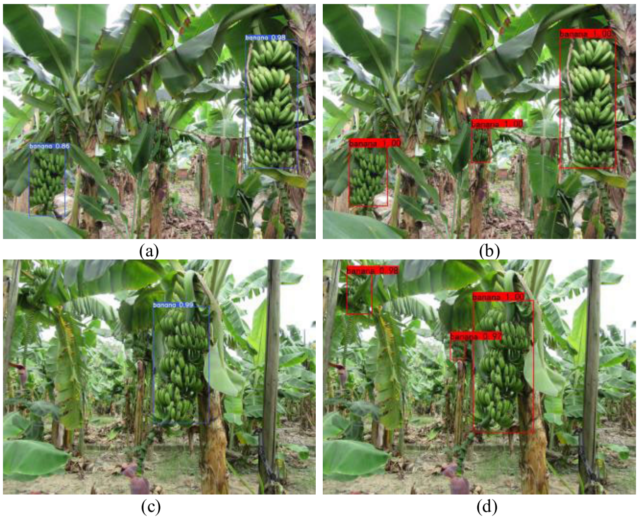


FIGURE 13. Detection results of the three hands of bananas: (a) YOLOv3; (b) YOLOv4; (c) the detection result of the small size of bananas based on YOLOv3; (d) the detection result of the small size of bananas based on YOLOv4.

judged the small banana as the background. Similarly, in the detection of three hands of bananas, as shown in Fig. 13, YOLOv3 misjudged the small size of bananas. The contrast is especially obvious in Fig. 13 (c) (d), for the left two hands of bananas, the human eye may have to distinguish carefully to see the location of the fruit. YOLOv3 misjudged the fruit, whereas YOLOv4 made an accurate detection. This is due to the innovation of the structure and the use of tricks.

Small object fruit detection is of great significance to the production management of banana plantations. First, different varieties of bananas vary in size, and small object detection can reasonably judge different varieties; Moreover, the accurate detection of small fruit can provide useful information for continuous detection.

Finally, we tried to detect the banana captured at an elevation angle. Although most bananas can be captured horizontally, some banana plants are still tall and the fruit is very high from the ground. We conducted experiments on the banana images with the angle of elevation to see whether the banana can be detected. Fig.14 shows the detection results of YOLOv3 and YOLOv4 in the elevation angle image. It can be seen that YOLOv4 accurately detected the banana fruits, but YOLOv3 failed to detect them. YOLOv4 has better generalization ability.

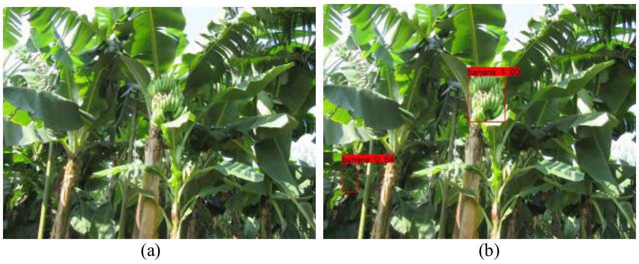


FIGURE 14. Comparison of detection results taken at elevation angle; (a) YOLOv3; (b) YOLOv4.

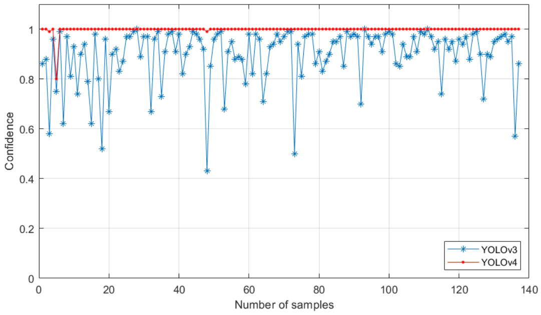


FIGURE 15. Comparison of confidence between the two detection algorithms.

From the above comparison, it is noticed that the confidence of the two methods was different. Therefore, we compared the confidence levels of the bananas detected by the two algorithms in 120 images. The results are shown in FIGURE. The detection confidence of YOLOv3 was between 0.5 and 1.0, whereas that of YOLOv4 was almost 1.0. When the banana was covered more area, the confidence level would be low.

B. COMPARISON BETWEEN DEEP LEARNING ALGORITHM AND MACHINE LEARNING ALGORITHM

We compared the banana detection results of YOLOv4, YOLOv3, and HOG+LBP+SVM algorithms. The detection results of the three algorithms in different conditions have been described in detail in the above and literature [60], The problems encountered in the machine learning algorithm are described below. In literature [60], when the key parts of the banana are covered, the banana was mistaken as two hands of bananas. We carried out experiments in YOLOv3 and

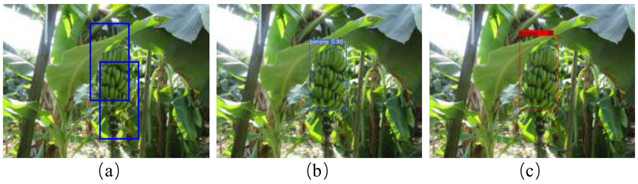

(a) (b) (c)

FIGURE 16. The detection results of the three algorithms under occlusion conditions: (a) HOG+LBP+SVM; (b) YOLOv3; (c) YOLOv4.

YOLOv4. As shown in Fig. 16, YOLOv3 detected the banana with a confidence of 0.90, and the fruit area at the top of the banana was not completely detected, whereas the result of YOLOv4 is more accurate.

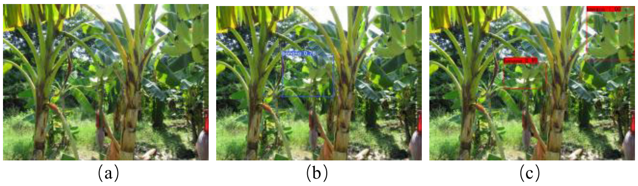

(a) (b) (c)

FIGURE 17. Detection results of the three algorithms for MUSA AA banana: (a) HOG+LBP+SVM; (b) YOLOv3; (a) YOLOv4.

MUSA AA banana detection results using the three algorithms were compared, as is shown in Fig. 17. Because of the poor growth, the banana had very few fingers whereas the illumination was strong, the contrast between fruit and leaves was very small. Due to the limitation of the sliding window scale in the machine learning algorithm, the detection failed. YOLOv3 detected a hand of banana but lost the upper right hand of banana. For YOLOv4, the two hands of banana had been detected.

We analyzed the detection of banana from traditional machine learning algorithm to deep learning algorithm. In different conditions, the three algorithms can realize banana detection in terms of their respective detection capabilities. Table 2 compares the key indexes of the three algorithms in the same test set. It can be found that, compared with the deep learning algorithm, the machine learning algorithm has a lower running cost, shorter training time, and smaller weight file size, and it can be implemented on the CPU, does not need GPU, but longer detection time and lower detection rate. In deep learning algorithms, both YOLOv3 and YOLOv4 require GPU. To obtain the optimal model, YOLOv3 required 300 epochs training, whereas YOLOv4 needed 150 epochs. The training time of YOLOv3 was shorter than that of YOLOv4, but the weight file was larger than that of YOLOv4. YOLOv3 had the shortest detection time for a single image and a higher detection rate than machine learning, but it was not as high as YOLOv4. The detection rate of YOLOv4 was 99.29%, which was far higher than the other two algorithms. The average detection time of YOLOv4 was 0.171s and the shortest detection time was 0.135s. Since the network of YOLOv4 is deeper than that of YOLOv3, the detection time was also increased. On the whole, YOLOv4 could obtain the optimal weight model with fewer iterations in the training stage, and superior to the traditional machine learning algorithm and YOLOv3 algorithm with its high confidence and high detection rate the detection stage.

TABLE 2. Detection indexes of the three algorithms.

Algorithm	HOG+LBP +SVM	YOLOv3	YOLOv4
Training time	2.35 h	6.32 h	12.5 h
Weight file size	15.5 MB	469 MB	244 MB
Hardware platform	Intel(R) Core (TM) i7 – 5500U @2.4 GHz, 16.0 GB RAM, NVIDIA GeForce 940M	Intel(R) Core (TM) i7 – 9750H @2.6 GHz 2.59GHz, 16.0 GB RAM, NVIDIA GeForce RTX 2070 with Max-Q Design	Intel(R) Core (TM) i7 – 9750H @2.6 GHz 2.59GHz, 16.0 GB RAM, NVIDIA GeForce RTX 2070 with Max-Q Design
Test set	120 images	120 images	120 images
The average detection time	1.325s	0.038s	0.171s
The shortest detection time	0.343s	0.030s	0.135s
Detection rate	89.63%	90.78%	99.29%

VI. CONCLUSION

The accurate detection of banana is of great significance to the intelligent management of the banana plantation. In this article, we proposed a detection method based on the latest YOLOv4 neural network for the banana detection in the natural environment. Besides, we analyzed the performance of the traditional machine learning algorithm and another neural network algorithm in banana detection. According to the experimental results, the following conclusions can be summarized:

(1) We found the suitable deep learning algorithm for banana detection in the plantation. The structural characteristics of the YOLOv4 neural network and the key problems of banana detection were analyzed. In the network, CSPDarknet53 deepens the network, which could extract more deep banana features and reduce the interference of green background to the green and irregular banana fruit; The SPP structure increases the acceptance range of network features with less computational cost, FPN+PAN structure repeatedly fuses multi-scale features to extract more profound banana semantic information and positioning information, to detect more precise banana fruits. Precise detection can still be achieved when the sizes of the banana fruits in the same image are greatly different; The *DIOU_nms* algorithm improves the confidence of banana detection results.

(2) The banana detection algorithm in the plantation based on YOLOv4 can achieve accurate detection under the

conditions of different illumination and occlusion for different varieties and maturity, providing precise information for the banana plantation intelligent management and fruit picking.

(3) The detection performance of deep learning algorithm is better than that of machine learning algorithm for banana detection in the plantation. Compared with HOG+LBP+SVM, YOLOv3, and YOLOv4, the average detection time of the three algorithms was 1.325s, 0.038s, and 0.171s. The detection rate of banana was 89.63%, 90.78%, and 99.29%, respectively. In the training stage, YOLOv4 obtained the optimal weight model with fewer iterations. In the detection stage, YOLOv4 was superior to the traditional machine learning algorithm and YOLOv3 algorithm with its high confidence and high detection rate.

In conclusion, the proposed method is suitable for banana detection in the plantation. The future work will be mainly to obtain the coordinate value of banana fruit in the real world, realize the localization of banana fruit, and calculate the location of the picking point.

REFERENCES

[1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' 2020, *arXiv:2004.10934*. [Online]. Available: http://arxiv.org/abs/2004.10934

[2] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, *arXiv: 1804.2767*. [Online]. Available: https://arxiv.org/abs/1804.2767

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Oct. 1998.

[4] K. Alex, S. Ilya, and E. H. Geoffrey, ''ImageNet Classification with deep convolutional neural networks,'' in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[5] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[6] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' 2015, *arXiv: 1512.3385*. [Online]. Available: https://arxiv.org/abs/1512.3385

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ''Going deeper with convolutions,'' 2014, *arXiv:1409.4842*. [Online]. Available: http://arxiv.org/abs/1409.4842

[8] S. Ioffe and C. Szegedy, ''Batch normalization: Accelerating deep network training by reducing internal covariate shift,'' 2015, arXiv: 1502.3167. [Online]. Available: https://arxiv.org/abs/1502.3167

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' 2015, *arXiv: 1512.1567*. [Online]. Available: https://arxiv.org/abs/1512.1567

[10] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, ''Inception-v4, inception-resNet and the impact of residual connections on learning,'' 2016, *arXiv: 1602.7261*. [Online]. Available: https://arxiv.org/abs/1602.7261

[11] F. Chollet, ''Xception: Deep learning with depthwise separable convolutions,'' 2016, *arXiv: 1610.2357*. [Online]. Available: https://arxiv.org/abs/1610.2357

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' 2013, *arXiv:1311.2524*. [Online]. Available: http://arxiv.org/abs/1311.2524

[13] R. Girshick, ''Fast R-CNN,'' 2015, *arXiv: 1504.8083*. [Online]. Available: https://arxiv.org/abs/1504.8083

[14] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards real-time object detection with region proposal networks,'' 2015, arXiv: 1497.1506. [Online]. Available: https://arxiv.org/abs/1497.1506

[15] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' 2016, *arXiv: 1612.3144*. [Online]. Available: https://arxiv.org/abs/1612.3144

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' 2015, *arXiv: 1506.2640*. [Online]. Available: https://arxiv.org/abs/1506.2640

[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, ''SSD: Single shot multiBox detector,'' 2015, *arXiv: 1512.2325*. [Online]. Available: https://arxiv.org/abs/1512.2325

[18] J. Redmon and A. Farhadi, ''YOLO9000: Better, faster, stronger,'' 2016, *arXiv: 1612.8242*. [Online]. Available: https://arxiv.org/abs/1612.8242

[19] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' 2014, *arXiv:1411.4038*. [Online]. Available: http://arxiv.org/abs/1411.4038

[20] V. Badrinarayanan, A. Kendall, and R. Cipolla, ''SegNet: A deep convolutional encoder-decoder architecture for image segmentation,'' 2015, *arXiv: 1511.1561*. [Online]. Available: https://arxiv.org/abs/1511.1561

[21] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ''Semantic image segmentation with deep convolutional nets and fully connected CRFs,'' 2014, *arXiv:1412.7062*. [Online]. Available: http://arxiv.org/abs/1412.7062

[22] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ''DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,'' , vol. 2016, arXiv: 1606.1915. [Online]. Available: https://arxiv.org/abs/1606.1915

[23] L. Chen, G. Papandreou, F. Schroff, and H. Adam, ''Rethinking Atrous convolution for semantic image segmentation,'' 2017, *arXiv: 1706.5587*. [Online]. Available: https://arxiv.org/abs/1706.5587

[24] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, ''Encoder-decoder with Atrous separable convolution for semantic image segmentation,'' 2018, *arXiv: 1802.2611*. [Online]. Available: https://arxiv.org/abs/1802.2611

[25] C. Liu, L. Chen, F. Schroff, H. Adam, W. Hua, A. Yuille, and F. Li, ''Auto-deepLab: Hierarchical neural architecture search for semantic image segmentation,'' 2019, *arXiv: 1901.2985*. [Online]. Available: https://arxiv.org/abs/1901.2985

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, ''Mask R-CNN,'' 2017, *arXiv: 1703.6870*. [Online]. Available: https://arxiv.org/abs/1703.6870

[27] M. Chen, Y. Tang, X. Zou, K. Huang, L. Li, and Y. He, ''High-accuracy multi-camera reconstruction enhanced by adaptive point cloud correction algorithm,'' *Opt. Lasers Eng.*, vol. 122, pp. 170–183, Nov. 2019, doi: 10.1016/j.optlaseng.2019.06.011.

[28] Y. Tang, L. Li, C. Wang, M. Chen, W. Feng, X. Zou, and K. Huang, ''Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision,'' *Robot. Comput.-Integr. Manuf.*, vol. 59, pp. 36–46, Oct. 2019, doi: 10.1016/j.rcim.2019.03.001.

[29] Y. Tang, M. Chen, C. Wang, L. Luo, J. Li, G. Lian, and X. Zou, ''Recognition and localization methods for vision-based fruit picking robots: A review,'' *Frontiers Plant Sci.*, vol. 11, May 2020, doi: 10.3389/fpls.2020.00510.

[30] Y. Majeed, J. Zhang, X. Zhang, L. Fu, M. Karkee, Q. Zhang, and M. D. Whiting, ''Deep learning based segmentation for automated training of apple trees on trellis wires,'' *Comput. Electron. Agricult.*, vol. 170, Mar. 2020, Art. no. 105277, doi: 10.1016/j.compag.2020.105277.

[31] K. Bresilla, G. D. Perulli, A. Boini, B. Morandi, L. C. Grappadelli, and L. Manfrini, ''Single-shot convolution neural networks for real-time fruit detection within the tree,'' *Frontiers Plant Sci.*, vol. 10, p. 611, May 2019, doi: 10.3389/fpls.2019.00611.

[32] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng, ''Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot,'' *Comput. Electron. Agricult.*, vol. 172, May 2020, Art. no. 105380, doi: 10.1016/j.compag.2020.105380.

[33] H. Kang, H. Zhou, and C. Chen, ''Visual perception and modeling for autonomous apple harvesting,'' *IEEE Access*, vol. 8, pp. 62151–62163, 2020, doi: 10.1109/ACCESS.2020.2984556.

[34] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, ''Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of MangoYOLO,'' *Precis. Agricult.*, vol. 20, no. 6, pp. 1107–1135, Feb. 2019, doi: 10.1007/s11119-019-09642-0.

[35] R. Kestur, A. Meduri, and O. Narasipura, ''MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard,'' *Eng. Appl. Artif. Intell.*, vol. 77, pp. 59–69, Jan. 2019, doi: 10.1016/j.engappai.2018.09.011.

[36] R. Shi, T. Li, and Y. Yamaguchi, "An attribution-based pruning method for real-time mango detection with YOLO network," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105214, doi: 10.1016/j.compag.2020.105214.

[37] J. Xiong, Z. Liu, S. Chen, B. Liu, Z. Zheng, Z. Zhong, Z. Yang, and H. Peng, "Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method," *Biosyst. Eng.*, vol. 194, pp. 261–272, Jun. 2020, doi: 10.1016/j.biosystemseng.2020.04.006.

[38] C. Liang, J. Xiong, Z. Zheng, Z. Zhong, Z. Li, S. Chen, and Z. Yang, "A visual detection method for nighttime litchi fruits and fruiting stems," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105192, doi: 10.1016/j.compag.2019.105192.

[39] J. Li, Y. Tang, X. Zou, G. Lin, and H. Wang, "Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots," *IEEE Access*, vol. 8, pp. 117746–117758, Jul. 2020, doi: 10.1109/ACCESS.2020.3005386.

[40] Y. Majeed, M. Karkee, Q. Zhang, L. Fu, and M. D. Whiting, "Determining grapevine cordon shape for automated green shoot thinning using semantic segmentation-based deep learning networks," *Comput. Electron. Agricult.*, vol. 171, Apr. 2020, Art. no. 105308, doi: 10.1016/j.compag.2020.105308.

[41] T. T. Santos, L. L. de Souza, A. A. Dos Santos, and S. Avila, "Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association," *Comput. Electron. Agr.*, vol. 170, Jan. 2020, Art. no. 105247, doi: 10.1016/j.compag.2020.105247.

[42] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on mask-RCNN," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104846, doi: 10.1016/j.compag.2019.06.001.

[43] R. Kirk, G. Cielniak, and M. Mangan, "Lab fruits: A rapid and robust outdoor fruit detection system combining bio-inspired features with one-stage deep learning networks," *Sensors*, vol. 20, no. 1, p. 275, Jan. 2020, doi: 10.3390/s20010275.

[44] H. Xie, N. Dai, X. Yang, K. Zhan, and J. Liu, "Research on recognition methods of pomelo fruit hanging on tees base on machine vision," in *Proc. Annu. Int. Meeting*, Boston, MA, USA, Jul.2019, Art. no. 1900411, doi: 10.13031/aim.201900411.

[45] Z. Liu, J. Wu, L. Fu, Y. Majeed, Y. Feng, R. Li, and Y. Cui, "Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion," *IEEE Access*, vol. 8, pp. 2327–2336, 2020, doi: 10.1109/ACCESS.2019.2962513.

[46] Y. Wang, J. Lv, L. Xu, Y. Gu, L. Zou, and Z. Ma, "A segmentation method for waxberry image under orchard environment," *Scientia Horticulturae*, vol. 266, May 2020, Art. no. 109309, doi: 10.1016/j.scienta.2020.109309.

[47] G. Lin, Y. Tang, X. Zou, J. Xiong, and J. Li, "Guava detection and pose estimation using a low-cost RGB-D sensor in the field," *Sensors*, vol. 19, no. 2, p. 428, Jan. 2019, doi: 10.3390/s19020428.

[48] P. A. Dias, A. Tabb, and H. Medeiros, "Apple flower detection using deep convolutional networks," *Comput. Ind.*, vol. 99, pp. 17–28, Mar. 2018, doi: 10.1016/j.compind.2018.03.010.

[49] A. Koirala, K. B. Walsh, Z. Wang, and N. Anderson, "Deep Learning for Mango (Mangifera indica) Panicle Stage Classification," *Agronomy*, vol. 10, no. 1, p. 143, Jan. 2020, doi: 10.3390/agronomy10010143.

[50] P. Lin and Y. Chen, "Detection of strawberry flowers in outdoor field by deep neural network," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2018, pp. 482–486.

[51] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Instance segmentation of apple flowers using the improved mask R–CNN model," *Biosystems Eng.*, vol. 193, pp. 264–278, May 2020, doi: 10.1016/j.biosystemseng.2020.03.008.

[52] Z.-F. Xu, R.-S. Jia, Y.-B. Liu, C.-Y. Zhao, and H.-M. Sun, "Fast method of detecting tomatoes in a complex scene for picking robots," *IEEE Access*, vol. 8, pp. 55289–55299, 2020, doi: 10.1109/ACCESS.2020.2981823.

[53] M. Rahnemoonfar and C. Sheppard, "Real-time yield estimation based on deep learning," *SPIE*, Vol. 10218, Oct. 2017, Art. no. 1021809.

[54] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors*, vol. 20, no. 7, p. 2145, Apr. 2020, doi: 10.3390/s20072145.

[55] F. P. Boogaard, K. S. A. H. Rongen, and G. W. Kootstra, "Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging," *Biosystems Eng.*, vol. 192, pp. 117–132, Apr. 2020, doi: 10.1016/j.biosystemseng.2020.01.023.

[56] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "DeepFruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, no. 8, p. 1222, Aug. 2016, doi: 10.3390/s16081222.

[57] B. Neupane, T. Horanont, and N. D. Hung, "Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV)," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0223906, doi: 10.1371/journal.pone.0223906.

[58] A. Clark and J. McKechnie, "Detecting banana plantations in the wet tropics, australia, using aerial photography and U-Net," *Appl. Sci.*, vol. 10, no. 6, p. 2017, Mar. 2020, doi: 10.3390/app10062017.

[59] M. Chen, Y. Tang, X. Zou, K. Huang, Z. Huang, H. Zhou, C. Wang, and G. Lian, "Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology," *Comput. Electron. Agricult.*, vol. 174, Jul. 2020, Art. no. 105508, doi: 10.1016/j.compag.2020.105508.

[60] L. Fu, J. Duan, X. Zou, G. Lin, S. Song, B. Ji, and Z. Yang, "Banana detection based on color and texture features in the natural environment," *Comput. Electron. Agricult.*, vol. 167, Dec. 2019, Art. no. 105057, doi: 10.1016/j.compag.2019.105057.

LANHUI FU received the B.S. degree from Zhengzhou University, China, in 2009, and the M.S. degree from the Harbin Institute of Technology, China, in 2012. She is currently pursuing the Ph.D. degree with the College of Engineering, South China Agricultural University, China. Her research interests include machine vision and deep learning.

JIELI DUAN received the B.S. degree from Shanxi Agricultural University, China, in 1997, and the M.S. and Ph.D. degrees from South China Agricultural University, China, in 2008 and 2018, respectively. She is currently an Associate Professor with the College of Engineering, South China Agricultural University. Her research interests include agricultural machinery, agricultural engineering, and information.

XIANGJUN ZOU received the B.S. degree from Hunan Agricultural University, China, in 1981, and the Ph.D. degree from the Guangdong University of Technology, China, in 2005. She is currently a Professor with the College of Engineering, South China Agricultural University, China. Her research interests include virtual reality, agricultural robot, machine vision, and image processing.

JIAQUAN LIN received the B.S. degree from South China Agricultural University, China, in 2020, where he is currently pursuing the M.S. degree with the College of Engineering. His research interests include computer vision and deep learning.

LEI ZHAO received the B.S. degree from the Mianyang Teachers' College, China, in 2014, and the M.S. degree from Southwest University, China, in 2017. He is currently pursuing the Ph.D. degree with the College of Engineering, South China Agricultural University, China. His research interests include precision agriculture and automatic control.

JINHUI LI received the B.S. degree in computer science and technology from Hainan Normal University, China, in 2015, and the M.S. degree in computer application technology from South China Agricultural University, China, in 2018, where she is currently pursuing the Ph.D. degree with the College of Engineering. Her research interests include computer vision, image processing, and artificial intelligence.

ZHOU YANG received the B.S. and M.S. degrees from Shanxi Agricultural University, China, in 1995 and 1998, respectively, and the Ph.D. degree from South China Agricultural University, China, in 2001. He is currently a Professor with the College of Engineering, South China Agricultural University. He is also the President of Jiaying University, China. His research interests include mechanization of fruit production and intelligent agricultural machinery equipment. He is also a member of the Chinese Society of Agricultural Engineering.