agronomy

Article

YOLO-Banana: A Lightweight Neural Network for Rapid Detection of Banana Bunches and Stalks in the Natural Environment

Lanhui Fu [1], Zhou Yang [1,2], Fengyun Wu [1], Xiangjun Zou [1,3], Jiaquan Lin [1], Yongjun Cao [4,5,*] and Jieli Duan [1,*]

[1] College of Engineering, South China Agricultural University, Guangzhou 510642, China; lanhuifu2020@163.com (L.F.); yangzhou@scau.edu.cn (Z.Y.); fyseagull@163.com (F.W.); xjzou1@163.com (X.Z.); m15813345752@163.com (J.L.)

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

[3] Foshan-Zhongke Innovation Research Institute of Intelligent Agriculture and Robotics, Foshan 528010, China

[4] School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China

[5] Institute of Intelligent Manufacturing, GDAS, Guangzhou 510075, China

[*] Correspondence: cyjauto@163.com (Y.C.); duanjieli@scau.edu.cn (J.D.)

Abstract: The real-time detection of banana bunches and stalks in banana orchards is a key technology in the application of agricultural robots. The complex conditions of the orchard make accurate detection a difficult task, and the light weight of the deep learning network is an application trend. This study proposes and compares two improved YOLOv4 neural network detection models in a banana orchard. One is the YOLO-Banana detection model, which analyzes banana characteristics and network structure to prune the less important network layers; the other is the YOLO-Banana-l4 detection model, which, by adding a YOLO head layer to the pruned network structure, explores the impact of a four-scale prediction structure on the pruning network. The results show that YOLO-Banana and YOLO-Banana-l4 could reduce the network weight and shorten the detection time compared with YOLOv4. Furthermore, YOLO-Banana detection model has the best performance, with good detection accuracy for banana bunches and stalks in the natural environment. The average precision (AP) values of the YOLO-Banana detection model on banana bunches and stalks are 98.4% and 85.98%, and the mean average precision (mAP) of the detection model is 92.19%. The model weight is reduced from 244 to 137 MB, and the detection time is shortened from 44.96 to 35.33 ms. In short, the network is lightweight and has good real-time performance and application prospects in intelligent management and automatic harvesting in the banana orchard.

Keywords: banana detection; stalk detection; improved YOLOv4; green fruit; orchard

check for updates

Citation: Fu, L.; Yang, Z.; Wu, F.; Zou, X.; Lin, J.; Cao, Y.; Duan, J. YOLO-Banana: A Lightweight Neural Network for Rapid Detection of Banana Bunches and Stalks in the Natural Environment. *Agronomy* **2022**, *12*, 391. https://doi.org/10.3390/agronomy12020391

Academic Editor: Simon Pearson

Received: 1 January 2022
Accepted: 29 January 2022
Published: 4 February 2022

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

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

1. Introduction

There are more than 130 countries in the world that cultivate bananas. In 2020, the output of bananas in China was 11.113 million tons. Bananas are used as the main food in some tropical regions because they are rich in vitamin A and fiber, and have high nutritional value. At present, banana orchards are mainly managed by banana farmers. The harvesting of bananas in the orchard also essentially relies on human labor [1]. Some mechanized transportation equipment has been gradually put into use in the banana orchards, but it still lags behind other fruits and vegetables in banana orchards in the research of intelligent management and automatic picking. In the complex environment of banana orchards, fast and accurate detection of banana bunches and stalks based on vision is the key task for the intelligent management of banana orchards. It provides solutions for saving labor and time costs, meeting high-quality fruit requirements, and reducing statistical errors. Therefore, solving the problems caused by occlusion, uneven illumination, and other unpredictable factors in the natural environment is one of the tasks to achieve accurate detection [2]. At the same time, a real-time and lightweight detection algorithm is also the key to promoting the intelligent development of banana orchards.

In the past few decades, the research of machine vision in fruit and vegetable detection has been rapidly updated with the development of artificial intelligence technology [3]. Before the explosion of deep learning, most fruit and vegetable detection methods were based on traditional machine learning algorithms, which are mostly based on hand-designed features (color [4,5], shape [6,7], texture [8,9], or fusion features [10,11], etc.) and appropriate classifiers (Adboost [8], support vector machine [12], etc.) to locate the object region in the image [13]. However, these methods often lack universality and robustness.

With the popularity of big data and the rapid development of GPUs, the application of deep learning algorithms in visual detection is advancing by leaps and bounds. Deep learning networks extract deeper features and have stronger learning ability. The deep neural network structure is more complex than traditional machine learning algorithms, the extracted features are more abstract, and the detection results have better generalization capabilities. Deep learning methods strike a balance between accuracy and real-time operation. Since LeCun proposed LeNet in 1998 [14], deep learning neural networks have been gradually applied and promoted in object classification, image segmentation, and object detection. In 2012, the emergence of AlexNet [15] pushed deep learning to the fore. In a complex agricultural environment, deep neural networks provide useful tools for the detection of fruits and vegetables. The classification network VGGNet was applied to the detection of red dates [16] and kiwifruits [17], which improved detection performance by increasing depth; Resnet has a deeper network but lower parameters than VGGNet. The improved Resnet was discussed for apples [18], strawberries [19], waxberries [20], and banana stalks [21]. The segmentation network solves the problem of image segmentation at the pixel level, e.g., FCN and SegNet neural networks were compared in the detection of grapevine cordon shape [22], and the Deeplabv3 series was used for multiple lychee fruit-bearing branches [23] and banana stalk segmentation [24]. An improved FCN was presented to detect the fruit center of guavas [25]. Compared with classification networks and segmentation networks, there are more application examples of detection networks in fruit and vegetable detection. The RCNN series are classical two-stage detectors that are based on the candidate regions, which were exploited in the detection of tomatoes [26], banana plants [27], and flowers [28]. An improved sweet pepper detection network DCNN based on Faster RCNN was proposed [29]. MobileNet is a small and efficient CNN model that offers a compromise between accuracy and latency. The authors of [30] used MobileNet to detect Hass avocado, lemon, and apples compared with Faster RCNN. EfficientDet trades off the speed and accuracy of the neural network; it was used by [31] to reconstruct the 3D global mapping of the orchard. YOLO series, which will be introduced in the following section, were used for cucumber internode length [32], kiwifruits [33], grapefruits [34], grapes [35], banana bunches [36], and banana bunches and stalks [37]. At the same time, improved networks based on YOLO series, namely MangoYOLO [38], YOLO-Tomato [39], and YOLOMuskmelon [40], were proposed; beyond these, the improved YOLOv3 was applied to detect the banana inflorescence axis [41] and an improved YOLOv5 method was described for apple detection [42].

Aiming at practical applications, this study compares the proposed YOLO-Banana network and YOLO-Banana-l4 network to seek a faster lightweight detection structure, to achieve real-time detection of banana bunches and stalks in banana orchards, while maintaining accuracy. Compared with the previous work, the main aim of this study is to optimize the model structure, reduce the detection time, and reduce the weight file according to the growth characteristics of banana bunches and stalks and the specific environment of the banana orchard, which will help to develop solutions for fruit detection and yield estimation in the banana orchard.

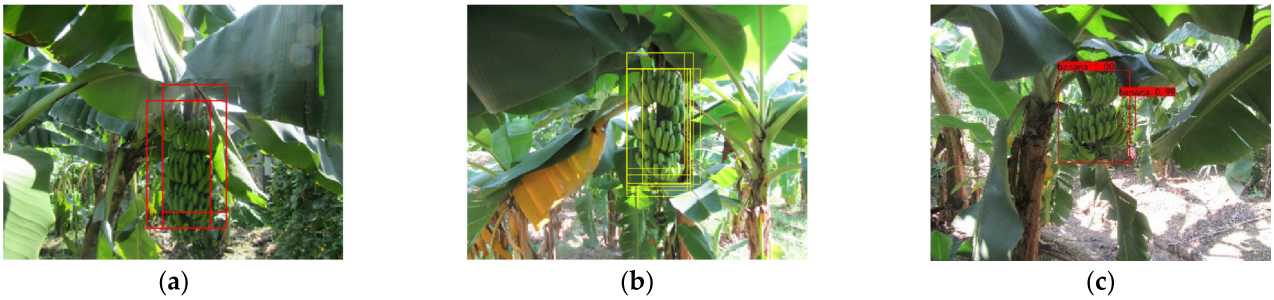
2. Materials and Methods

2.1. Image Acquisition

Considering the impact of the camera's angle of view and occlusion on the detection performance, banana images were collected from multiple angles during the image collection process. Under natural illumination conditions, images were collected from two banana orchards. In total, 388, 178, and 134 valid banana images were acquired at the banana plantation of Guangdong Academy of Agricultural Sciences on 9 August 2018 (sunny), 19 November 2018 (cloudy), and 16 March 2019 (overcast), and 464 valid banana images were acquired at Nansha banana plantation in Guangzhou on 27 October 2019 (sunny). This study focused on the detection of banana bunches and stalks in the complex natural environment, rather than distinguishing banana species. The capture device was a color digital camera (Canon sx610hs), the camera resolution was 2048 × 1536 pixels, the exposure mode was set to automatic exposure, the shooting distance was approximately 800~1200 mm when capturing, and the images were saved in JPG format. In the banana images, the banana bunches and stalks in the growth period were green, and the banana fingers pointed upward and curved in clusters. A labeling tool called Colabeler was used to label the banana bunches and stalks in the images. Each banana bunch and stalk in the images was manually labeled and checked twice to ensure accuracy. After the labeling was completed, an Extensible Markup Language (XML) file containing the information of the bunches and stalks and the position of the boundary rectangle was generated. Finally, we converted the label file into a txt file as the input. During training, the data set was randomly divided into training, validation, and test sets, whose sizes were 835, 209, and 120, respectively. In each image, there were usually one to three banana trees. Compared with the banana bunches, the bounding box of the banana stalks was much smaller. Since the banana stalk is connected to the pseudo-stem, and the texture is very close to the pseudo-stem, the detection of the stalks was more difficult than that of banana bunches. However, the detection of banana stalks is the key to intelligent harvesting, and the detection of banana bunches is an important indicator for growth management and yield estimation. Therefore, the accurate labeling of the banana bunches and stalks is an important prerequisite for accurate detection. We used Visual Studio 2019 to implement the algorithm on a laptop with Intel (R) Core (TM) i7—9750H @2.6 GHz 2.59 GHz, 16.0 GB RAM, NVIDIA GeForce RTX 2070 with Max-Q Design.

2.2. YOLO Series and Previous Work

YOLO series are one-stage detection algorithms and have a faster detection speed compared with two-stage algorithms such as R-CNN series, which are representative networks based on candidate regions. YOLO series solve object detection as a regression task, directly calculating the input image and outputting the class and corresponding positioning. The direct source of the YOLO series of algorithms is the sliding window technology, which converts the detection problem into an image classification problem. Our previous work [12] also performed single-scale and multi-scale detection of banana fruits based on the sliding window technology, as shown in Figure 1a,b. Since sliding windows of different sizes and proportions need to be set, a large amount of calculation is generated, and the detection time cost is high. A YOLO series algorithm changes the sliding window to directly divide the original image into non-overlapping grids. Each grid is responsible for the detection of the object whose center is in the grid, and predicts the bounding box of all the objects contained in the grid at one time, the confidence of location, and the probability vector of all categories. The YOLO series includes v1 to v5. The model has been improved in terms of input, network depth, backbone, neck, head, and output scale. The learning ability has been gradually enhanced, and the detection performance has been continuously improved. The model structure parameters are shown in Table 1. In the previous work [36], YOLOv4 was applied in banana orchards. As shown in Figure 1c, YOLOv4 has achieved accurate detection in banana orchards, especially for small objects.



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

Figure 1. Previous detection works in banana orchards: (**a**) single-scale detection, (**b**) multi-scale detection, (**c**) YOLOv4.

Table 1. Parameters of YOLO series.

Model	Input (Resolution)	Layers	Backbone	Neck	Bounding Boxes in Each Grid	Prediction (Resolution)
YOLOv1	448 × 448	24	Darknet19	–	2	13 × 13
YOLOv2	448 × 448	32	Darknet19	–	5	13 × 13
YOLOv3	416 × 416	106	Darknet53	FPN	9	13 × 13, 26 × 26, 52 × 52
YOLOv4	608 × 608	161	CSPDarknet53	FPN + PAN	9	19 × 19, 38 × 38, 76 × 76
YOLOv5	608 × 608	–	CSPDarknet53	FPN + PAN	9	19 × 19, 38 × 38, 76 × 76

2.3. Improvement Based on YOLOv4

The previous work [36] has verified that YOLOv4 has high detection accuracy in banana orchards. In this study, we aimed to enlarge the detection objects to banana stalks, and, at the same time, to improve the detection speed and make the network lightweight based on the previous results. In the detection of banana bunches and stalks, we simplified the network structure of YOLOv4, pruned unnecessary network structures, and removed redundant layers. The detection speed was finally improved on the basis of ensuring the detection accuracy, so as to realize the light weight of the network and the improvement of the detection speed, and to ensure the real-time performance.

This study proposes two improved models, the YOLO-Banana model and YOLO-Banana-l4 model, as shown in Figures 2 and 3. In the figures, the yellow module CBM (Convolutional + Batch Normalization + Mish) represents the convolution operation of Batch Normalization and the Mish activation function. The Mish activation function is used in the backbone part of the network to improve the accuracy of the network; in other parts of the network, the activation function still chooses the traditional Leaky ReLu function. The green module CBL (Convolutional + Batch Normalization + Leaky ReLu) represents the convolution operation of Batch Normalization and the Leaky ReLu activation function. The blue module represents the CSP structure, and CSPn contains n Res units, as shown in Figure 4. The orange module is the Concat operation, which corresponds to the route operation in the config file, which represents tensor splicing and expands the dimensions of the two tensors. The purple module represents the up-sampling operation, which implements FPN through up-sampling and then implements PAN through down-sampling through convolution operation to form a neck structure. The YOLO-Banana model keeps the backbone structure of YOLOv4, removes the SPP module, and simplifies the convolution numbers in the neck structure. In the YOLO-Banana model, the backbone is the CSPDarknet53 model, the neck structure is a FPN + PAN structure with 11 CBL modules, and the head part is the classic YOLO prediction part. In the head structure, the YOLO-Banana model has three scales. Each scale predicts three anchor boxes, with 7 values per anchor [4 box coordinates + 1 object confidence + 2 class confidences]. Therefore, the dimension of each scale is 21.

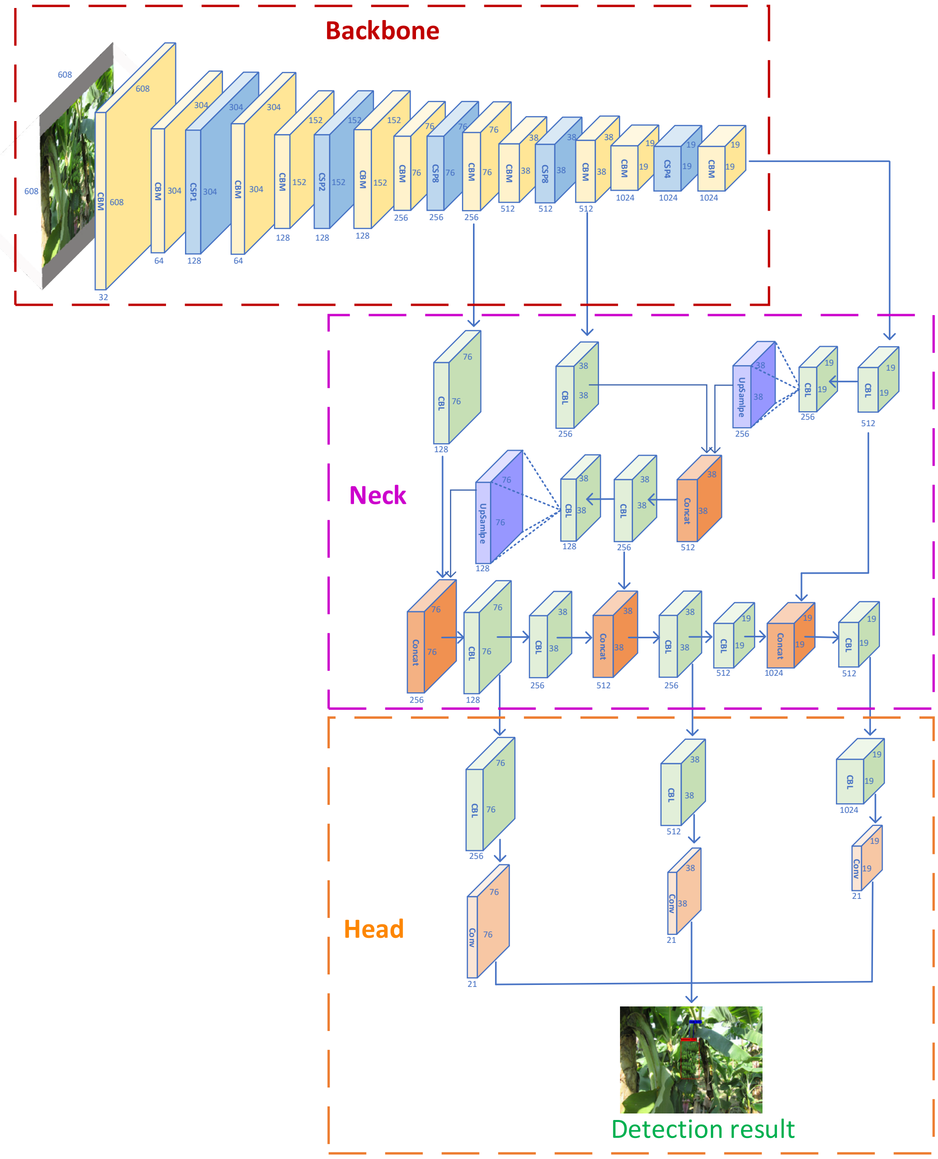


Figure 2. The detection flowchart of banana bunches and stalks based on YOLO-Banana.

On the basis of YOLO-Banana, the feature map generated by the second residual module in the backbone network was extracted. The size of the feature map was 152 × 152, and the tensor and dimension were merged with the third up-sampled feature map in the neck structure to convey the target information in more detail; then, the fused feature map was down-sampled through the convolution operation. Thereby adding a head layer, the head part became 4 layers, so it was called the YOLO-Banana-l4 model. The YOLO-Banana-l4 model has four scales in the head part, and each scale has 21 dimensions. The purpose was to observe the impact on detection performance by extracting a larger range of image features. Figure 3 describes the dimensions and meaning of the 4-layer layers of the YOLO-Banana-l4 model.

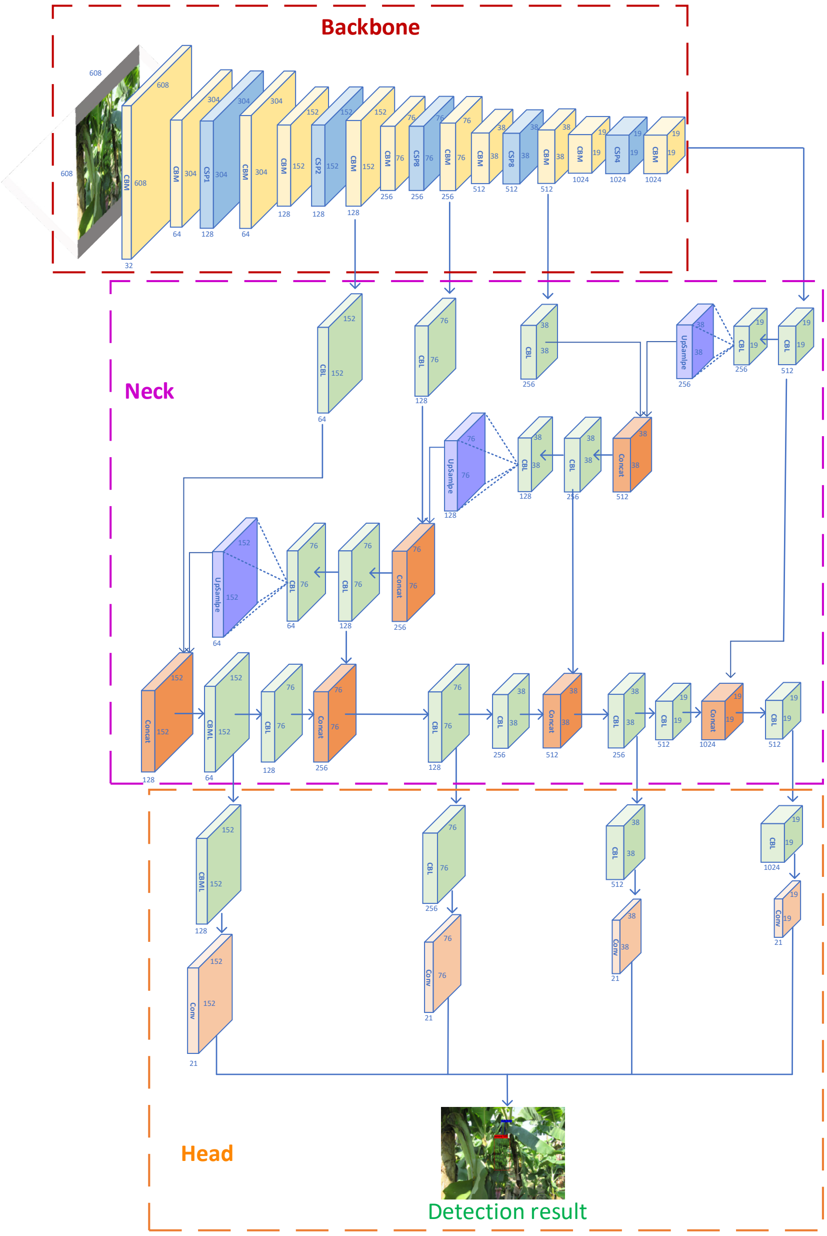


Figure 3. The detection flowchart of banana bunches and stalks based on YOLO-Banana-l4.

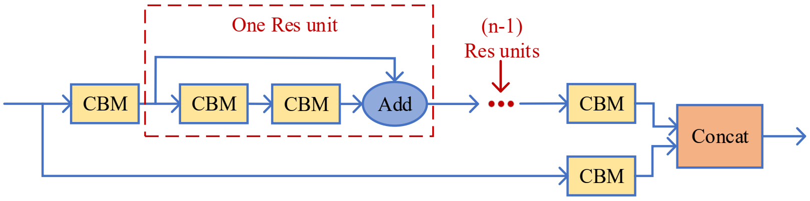


Figure 4. CSPn structure.

During training, all models use the same data set and hyperparameters. According to YOLO official advice, this study set the maximum number of iterations to 6000, batch to 64, subdivision to 16, and momentum to 0.949. The decay was set to 0.0005, and the initial value of the learning rate was set to 0.001. The steps were 3200 and 3600, and the corresponding scales were 0.1 and 0.1. If there was an object in the grid, only the bounding box with the largest IOU of ground truth was selected to be responsible for predicting the target, and other bounding boxes considered that there was no object. According to the class information predicted by each grid and the confidence information predicted by the bounding box, the class confidence of each box was calculated, and the score included an estimate of the accuracy of the class and the box, traversing the results of candidate boxes of all grids. The boxes whose confidence was less than the threshold were filtered out firstly, and then all boxes were sorted according to the confidence, and the bounding boxes with low scores were further filtered and removed by DIOU_nms, thereby obtaining a class of detection results. The banana bunches and stalks were processed to obtain the detection results of each class.

3. Results and Discussion

3.1. Model Evaluation

As the model structure had been changed, the official pre-weights of YOLOv4 were no longer applicable. Therefore, the YOLO-Banana model and YOLO-Banana-l4 model started training without pre-weighting, and after every 100 iterations, they were saved and updated to the latest weight file to use as pre-weights when training resumed after a training interruption. At the same time, the weight file of each 1000 iterations was saved as the training result, which was used to compare and analyze the training process of the YOLO-Banana model, YOLO-Banana-l4 model, and YOLOv4 model. The input resolution of the three models was 608 × 608, and the number of iterations was set to 6000. The training time is shown in Table 2. Compared with YOLOv4, YOLO-Banana's pruning of the model shortens the training time, and the four-layer head structure in YOLO-Banana-l4 increases the training time under the same input and number of iterations.

Table 2. Training time of different models.

Model	Iteration	Input (Resolution)	Training Time (h)
YOLOv4	6000	608 × 608	26.6
YOLO-Banana	6000	608 × 608	21.2
YOLO-Banana-l4	6000	608 × 608	24.1

The training loss curves of the three detection models are shown in Figure 5. Compared with YOLOv4, which started to converge after 330 iterations, YOLO-Banana started to converge after 400 iterations, while YOLO-Banana-l4 added a head layer so that the model could extract object features earlier and converged after 150 iterations. The loss of the three models gradually stabilized after 3500 iterations. It can be seen from the figure that the loss of YOLO-Banana is higher than the loss of YOLOv4 before stabilization, and the loss after stabilization is between YOLOv4 and YOLO-Banana-l4; although YOLO-Banana-l4 converges early, the decline rate of loss is the slowest, and the loss is higher than that of YOLO-Banana after 1300 iterations. From the trends of the convergence curve, the three models learned the object features well, and all the loss values after stabilization were less than 1, which shows that the models can be used in detection, similar to the literature [33].

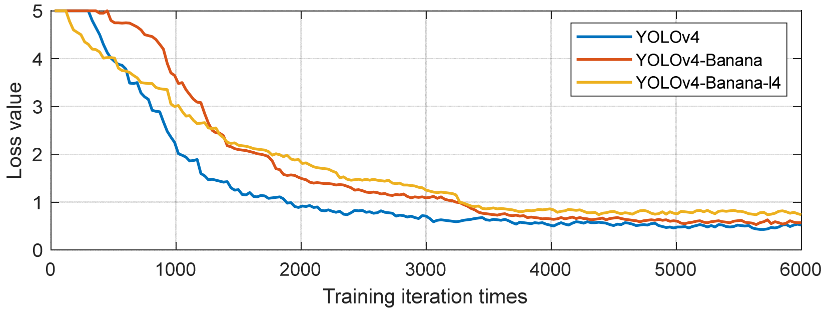


Figure 5. Training loss curves of different models.

We evaluate the training results and compare the AP values of the three detection models on banana bunches and stalks, and the mAP values of the entire model for all detection classes [33]. The two calculation formulas are (1) and (2):

$$AP = \int_0^1 P(R)dR \tag{1}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{2}$$

where P and R refer to the precision and recall of the detection model, respectively, and the calculation formulas are (3) and (4):

$$P = \frac{TP}{TP + FP} \times 100\% \tag{3}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

Among them, TP, FP, and FN are the abbreviations of True Positive, False Positive, and False Negative.

The YOLO-Banana model, YOLO-Banana-l4 model, and YOLOv4 model were used to verify the detection of banana bunches and stalks in the validation set. The evaluation results are shown in Table 3. As can be seen from the table, the AP values of YOLO-Banana for banana bunches and stalks are 98.4% and 85.98%, respectively, which are 1.16% and 2.1% lower than those of YOLOv4 (99.55% and 87.82%), and the mAP of YOLO-Banana (92.19%) is 0.84% lower than that of YOLOv4 (93.69%). The AP values of YOLO-Banana-l4 for banana bunches and stalks are 96.84% and 82.68%, which are 2.72% and 5.85% lower than those of YOLOv4. The mAP of YOLO-Banana-l4 model is 89.76%, which is 4.19% lower than that of YOLOv4. At the same time, the AP value of the banana bunches detected by the three detection models is significantly higher than that of the stalks. This is because the stalk size is much smaller and the texture is closer to the petiole, compared to banana bunches, so it is more difficult to detect banana stalks. Comparing the number of layers and weight file sizes of the three models, the YOLO-Banana network has a depth of 134 layers and the weight file is 137 MB, and the YOLO-Banana-l4 network has a depth of 147 layers and the weight file is 138 MB. The two models reduced the number of layers and reduced the model weight by nearly half compared to YOLOv4 (161 layers and 244 MB).

Table 3. Performance comparison of the model with different models.

Model	**AP**		**mAP (%)**	**Layer**	**Weight (MB)**
	Banana (%)	**Stalk (%)**			
YOLOv4	99.55	87.82	93.69	161	244
YOLO-Banana	98.4	85.98	92.19	134	137
YOLO-Banana-l4	96.84	82.68	89.76	147	138

Based on the above results and overall analysis, the YOLO-Banana model can save training time and reduce model weight while ensuring detection accuracy; the YOLO-Banana-l4 model realizes weight reduction, but, due to the addition of a head layer, the training time is not significantly reduced, and the AP value of the stalks is reduced, which affects the detection accuracy of the entire model. We further discuss the detection results of the two improved models in the test set and compare them with the YOLOv4 model to analyze the most suitable detection model in the banana orchard.

3.2. Detection Results under Different Illumination

We detected banana bunches and stalks under different illumination conditions, compared the two improved models with YOLOv4, and analyzed the detection performance of the three models.

There were 163 banana bunches and 141 stalks in the 120 banana images in the test set. In sunny conditions, including sunny front-light and sunny backlight environments, 62 images contained 80 banana bunches and 65 stalks; in cloudy conditions, 58 images contained 83 banana bunches and 76 stalks. YOLO-Banana, YOLO-Banana-l4, and YOLOv4 were applied in sunny and cloudy environments, and the numbers of correctly detected, falsely detected, and missed objects were counted, as shown in Table 4. The illumination of sunny conditions is higher than that of cloudy days, and the object features are easier to capture, but excessively bright light will lead to features becoming blurred. At the same time, the illumination of the sunny front-light and sunny backlight environment is also different. The illumination of cloudy conditions is more uniform than that of sunny conditions, but insufficient brightness also makes detection difficult. Therefore, whether a model can achieve robust detection under different illumination conditions is an important indicator to measure the quality of the model. It can be seen from the table that the detection results of banana bunches and stalks for the three models are generally good, regardless of whether it is sunny or cloudy. The detection accuracy of banana bunches is generally higher than that of stalks. This is related to the difference in size and texture. In contrast, the detection results of YOLO-Banana and YOLOv4 are close, slightly lower than YOLOv4; the detection results of YOLO-Banana-l4 are essentially the lowest. Comparing the detection results of banana bunches and stalks in sunny and cloudy conditions, respectively, the correct detection rate is very close, indicating that the improved models are robust to changes in illumination. It is easy to find that the missed detection rate of banana bunches and stalks is higher than the false detection rate. This is because the small-sized banana bunches or small-sized stalks are affected by occlusion, which easily leads to missed detection. The occlusion problem will be explained in further detail below. When the size of the banana bunches and stalks was large, falsely detected cases were rare. The performance of the three models was different. For example, when the stalk and the petiole were close, YOLO-Banana-l4 misjudged the petiole as the stalk, as shown in Figure 6. Regarding missed detection, YOLO-Banana-l4 has the highest missed detection rate for the stalks. Let us give an example to illustrate. For instance, in the sunny conditions shown in Figure 7 and the cloudy conditions in Figure 8. YOLOv4 and YOLO-Banana could detect banana bunches and stalks correctly, and YOLO-Banana-l4 could accurately detect banana bunches, but the stalks were missed, which occurred more frequently in the detection of small-sized stalks. In the Pisang Mas Musa (AA Group) banana detection results with strong light and fewer banana fingers, the missed detection rate of the three models was higher, as shown in Figure 9. Since the banana finger was very similar to the background, it was difficult to detect, even with the human eye. In the detection of the three models, YOLOv4 missed the stalk on the left, YOLO-Banana missed the fruit stalk on the right, and YOLO- Banana-l4 only detected one banana bunch and missed the right-hand banana bunch and two stalks.

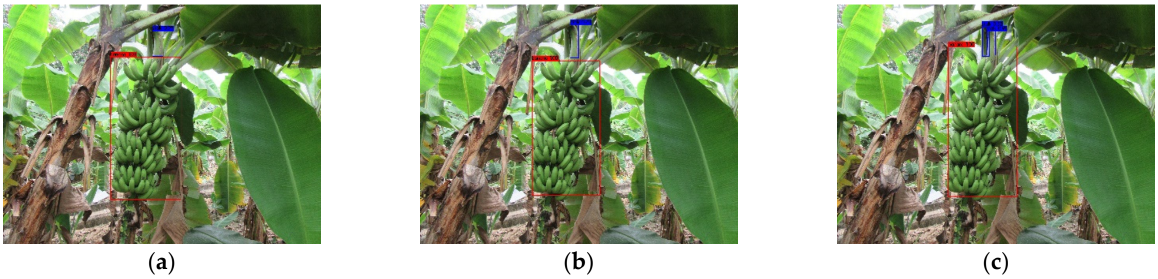


Figure 6. The detection result of the close interference between the petiole and stalk: (**a**) YOLOv4; (**b**) YOLO-Banana; (**c**) YOLO-Banana-l4.

Table 4. The detection results of the three methods under different illuminations.

Illumination	Object	Model	Count	Correctly Detected		Falsely Detected		Missed	
				Amount	Rate (%)	Amount	Rate (%)	Amount	Rate (%)
Sunny	Banana	YOLOv4	80	79	98.75	0	0	1	1.25
		YOLO-Banana	80	79	98.75	0	0	1	1.25
		YOLO-Banana-l4	80	78	97.5	0	0	2	2.5
	Stalk	YOLOv4	65	59	90.77	1	1.54	6	9.23
		YOLO-Banana	65	58	89.23	2	3.08	7	10.77
		YOLO-Banana-l4	65	57	87.69	1	1.54	8	12.31
Cloudy	Banana	YOLOv4	83	83	100	0	0	0	0
		YOLO-Banana	83	82	98.8	0	0	1	1.2
		YOLO-Banana-l4	83	79	95.18	0	0	4	4.82
	Stalk	YOLOv4	76	69	90.79	1	1.32	7	9.21
		YOLO-Banana	76	69	90.79	2	2.63	7	9.21
		YOLO-Banana-l4	76	65	85.53	4	5.26	11	14.74

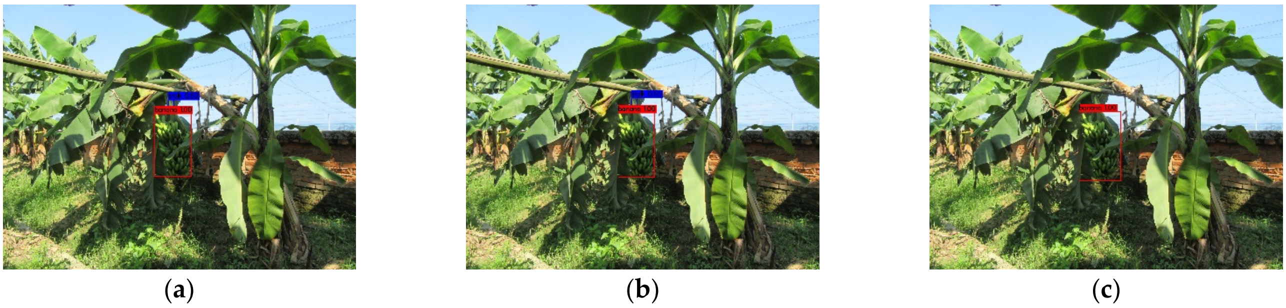


Figure 7. Comparison of the detection of the three models under sunny conditions: (**a**) YOLOv4 accurately detected the banana bunch and the stalk; (**b**) YOLO-Banana accurately detected the banana bunch and the stalk; (**c**) YOLO-Banana-l4 accurately detected the banana bunch, but the stalk was missed.

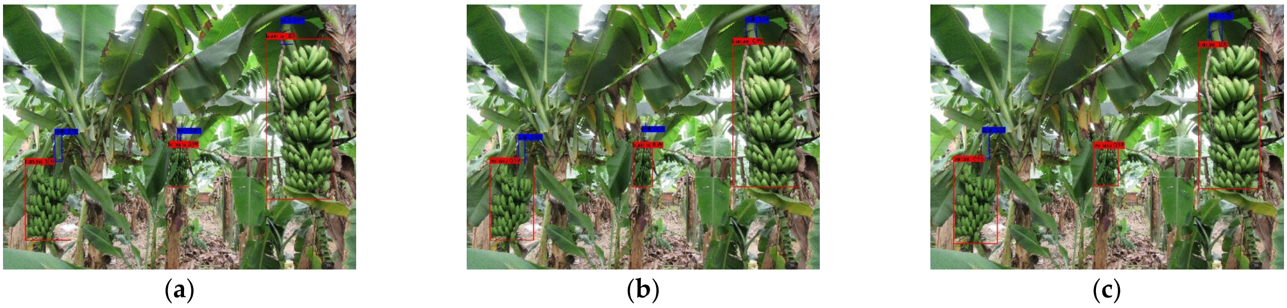
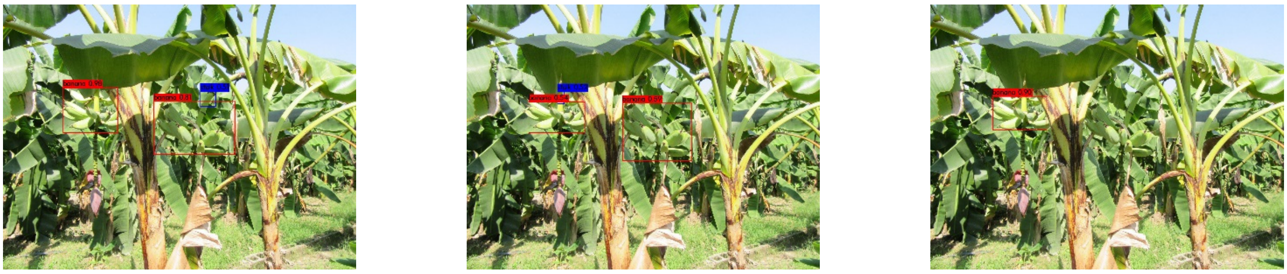


Figure 8. Comparison of the detection of the three models under cloudy conditions, (**a**) YOLOv4 accurately detected the banana bunch and the stalk; (**b**) YOLO-Banana accurately detected the banana bunch and the stalk; (**c**) YOLO-Banana-l4 accurately detected the banana bunch, but missed the middle stalk.



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

Figure 9. Examples of missed detection of the three models: (**a**) YOLOv4 missed the left stalk; (**b**) YOLO-Banana missed the right talk; (**c**) YOLO-Banana-l4 missed the left banana bunch and the right bunch with its stalk.

3.3. Detection Results under Different Occlusion Conditions

Banana bunches at close distances are larger in size, and the occluded area ratio is generally not too high, while bunches at far distances have a smaller field of view and are easily occluded by other banana bunches, branches, or dead leaves in front, and the occluded area ratio is also high. The size of the stalk is small, the growth position is high, and the cover of dead leaves is also common. In order to evaluate the detection performance of the improved model under different occlusion conditions, it was divided into slight occlusion and severe occlusion according to the degree of occlusion. In reality, considering the important role of short-range targets in production, when the occlusion area exceeds 20% of the banana bunch or stalk, it is considered to be a serious occlusion situation. In the test set, there were 50 banana bunches and 46 stalks in the case of slight occlusion, and 19 banana bunches and 13 stalks in the case of severe occlusion. If the occlusion is too serious, the banana bunches and stalks that are difficult to see even by the human eye will be regarded as the background, which has no semantic information meaning for the model. We tested the three models under different occlusion conditions and counted the results of correctly detected, falsely detected, and missed objects, as shown in Table 5.

Table 5. The detection results of the three methods under different occlusion conditions.

Occlusion	Object	Model	Count	Correctly Detected		Falsely Detected		Missed	
				Amount	Rate (%)	Amount	Rate (%)	Amount	Rate (%)
Slight	Banana	YOLOv4	50	50	100	0	0	0	0
		YOLO-Banana	50	50	100	0	0	0	0
		YOLO-Banana-l4	50	50	100	0	0	0	0
	Stalk	YOLOv4	46	40	86.96	1	2.17	3	6.52
		YOLO-Banana	46	42	91.3	2	4.35	4	8.7
		YOLO-Banana-l4	46	39	84.78	2	4.35	5	10.87
Severe	Banana	YOLOv4	19	18	94.74	0	0	1	5.26
		YOLO-Banana	19	17	89.47	0	0	2	10.53
		YOLO-Banana-l4	19	13	68.42	0	0	6	31.58
	Stalk	YOLOv4	13	10	76.92	0	0	6	46.15
		YOLO-Banana	13	9	69.23	0	0	4	30.77
		YOLO-Banana-l4	13	8	61.53	0	0	6	46.15

It can be found from the table that the three models showed high detection capabilities for banana bunches when slightly occluded, and the detection results of the stalks were obviously not as high as the correct rate of banana bunch detection. Among them, the correct rate of YOLO-Banana was the highest, followed by YOLOv4, and finally YOLO-Banana-l4. When severely occluded, YOLOv4 had the highest accuracy in detecting bunches and stalks, followed by YOLO-Banana, and finally YOLO-Banana-l4. Similarly, banana bunches had a higher accuracy rate than stalks under severe occlusion conditions. We illustrate three examples of different model detection results under slight occlusion and severe occlusion conditions. As shown in Figures 10 and 11, YOLO-Banana could achieve the same detection effect as YOLOv4 at both degrees, while YOLO-Banana-l4 model was prone to missed detection. It should be noted here that when the occlusion was severe, the false detection rates of the three models were all 0. This is because the far-distance banana bunches and stalks are more likely to lead to missed detection when the occlusion area increases. As shown in Figure 12, the small-sized banana bunch and stalk on the left-hand side of the figure were seriously blocked by branches and leaves. YOLOv4 detected the stalk and missed the banana bunch; YOLO-Banana successfully detected the fruit and fruit shaft, while YOLO-Banana-l4 missed the left stalk and only detected a part of the banana bunch in the left.

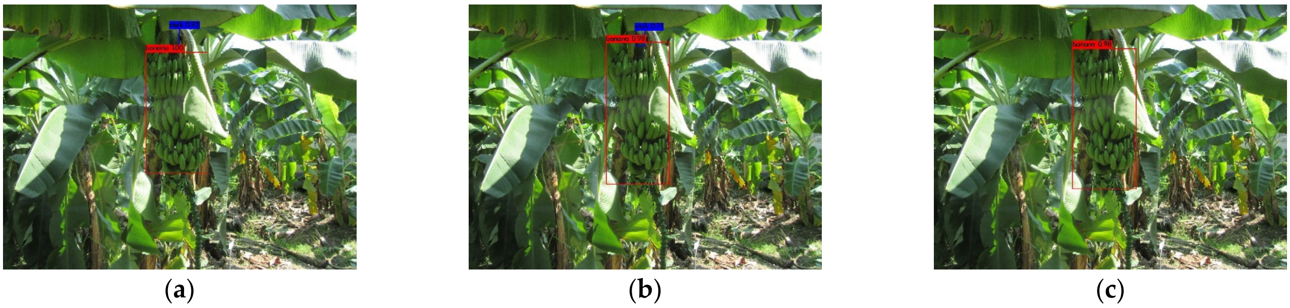

(**a**) (**b**) (**c**)

Figure 10. Detection results in slight occlusion conditions: (**a**) YOLOv4; (**b**) YOLO-Banana; (**c**) YOLO-Banana-l4.

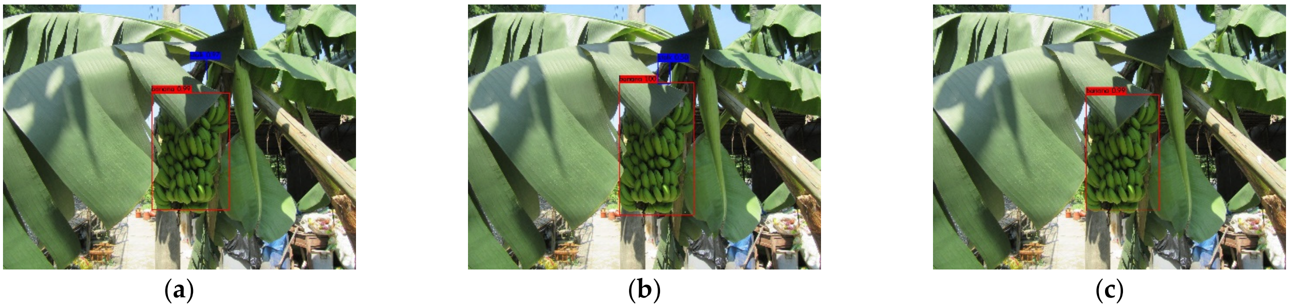

(**a**) (**b**) (**c**)

Figure 11. Detection results in severe occlusion conditions: (**a**) YOLOv4; (**b**) YOLO-Banana; (**c**) YOLO-Banana-l4.

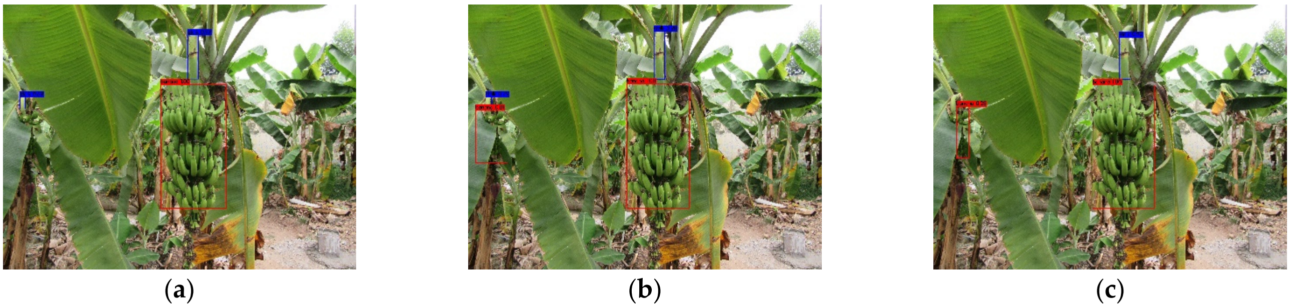

(**a**) (**b**) (**c**)

Figure 12. Detection results in the case of large area occlusion: (**a**) YOLOv4; (**b**) YOLO-Banana; (**c**) YOLO-Banana-l4.

From the analysis of different illumination and occlusion conditions, we can see that, in banana orchard detection, the detection accuracy rate of banana bunches is greater than that of stalks. Illumination has little effect on the detection model, while the degree of occlusion has a significant impact on it. The missed detection rate of banana bunches and stalks is higher than the false detection rate due to the influence of small-sized objects, occlusion, or banana variety. Regarding the detection results of the two improved models proposed in this study in different environments, the detection results of YOLO-Banana are similar to those of YOLOv4 and show better performance than YOLO-Banana-l4.

3.4. Confidence and Detection TIME

Finally, the average confidence and average detection time of the three models in the test set are compared, as shown in Table 6. YOLOv4 has the highest average detection confidence for banana bunches and stalks, respectively 0.96 and 0.91, followed by YOLO-Banana (0.94 and 0.89), and finally YOLO-Banana-l4 (0.92 and 0.88). It is worth mentioning that the detection confidence of small-sized objects is generally lower than that of close-distance objects. In terms of detection time, the average time for YOLO-Banana to detect a single image is 35.33ms, which is 21.42% less than YOLOv4 (44.96ms), and the average time for YOLO-Banana-l4 to detect a single image is 38.19 ms, which is 15.06% less than YOLOv4.

Table 6. Comparison of the confidence and the detection time of the three models.

Model	**Average Confidence**		**Time (ms)**
	Banana	**Stalk**	
YOLOv4	0.96	0.91	44.96
YOLO-Banana	0.94	0.89	35.33
YOLO-Banana-l4	0.92	0.88	38.19

According to the analysis of the results in the validation set and the test set, compared to the detection ability of the YOLOv4 model in banana gardens, the YOLO-Banana model can meet the high accuracy requirements of detection while reducing the layers and the weight of the model and saving the model training time and detection time. Compared with the YOLO-Banana model, the improvement of the YOLO-Banana-l4 model in terms of weight reduction and time saving is weaker. It also proves that the four-layer design does not significantly improve the detection effect of the YOLO series in banana orchards, and it also provides a reference for the improvement of YOLOv4 networks in other orchards. At present, the lightweight network structure is an application trend in intelligent agriculture. Without losing a large amount of recognition accuracy, the YOLO-Banana model has faster detection speed and takes up less storage space, which has practical significance for the intelligent detection of banana orchards.

3.5. Discussion

From the above results, the YOLO-Banana model reduces the detection time and model weight of YOLOv4 under the premise of ensuring accuracy. Compared with our previous works [12] and [36], this study extends the detection object to banana bunches and stalks and improves the performance of the detection model. We further compare the detection results of YOLOv4 and YOLO-Banana with other detection networks [37] and [41] for banana bunches and stalks, as shown in Table 7. The banana bunch detection rates of YOLOv4 and YOLO-Banana are higher than those in [37] and [41]. The banana stalk detection rates of YOLOv4 and YOLO-Banana are lower than that in [37], which is because we limit the banana stalk area to only contain the stalk, and [37] defines a much larger area for the stalks, which reduces the detection difficulty. In terms of detection time, the detection time in [41] is 240 ms, and [37] does not specify the detection time. It can be seen from the comparison that the model proposed in this study achieves good results in terms of detection accuracy and detection time, as well as the setting of the detection area.

Table 7. Comparison of the detection results by different models.

Model	Hardware Platform	AP		mAP (%)	Time (ms)
		Banana (%)	Stalk (%)		
YOLOv4	Intell CoITM) i7—9750H @2.6 GHz 2.59GHz, 16.0 GB RAM, NVIDIA GeForce RTX 2070 with Max-Q Design	99.55	87.82	93.69	44.96
YOLO-Banana		98.4	85.98	92.19	35.33
YOLOv3 [37]	2 GeForce RTX 2080 GPUIntIR) Xeon(R) CPU E5-2620 v4 @2.10GHz 2.10 GHz(2 processors)	88	98	93	unknown
Improved YOLOv3 [41]	i7-7700K processor, memory 16G,2,400 MHz; video card GTX1080Ti 11G	94	undetected	–	240

4. Conclusions

The real-time detection of banana bunches and stalks is an important part of the intelligent management and automatic harvesting of the banana orchard. In this study, two improved models, YOLO-Banana and YOLO-Banana-l4, are proposed based on YOLOv4. Through the comparative analysis of the results in the processes of training, verification, and testing, the following conclusions can be summarized. (1) We found a faster, lightweight detection model with 134 layers in the YOLO-Banana network model with a weight of 137MB. The AP values for banana bunch and stalk detection were 98.4% and 85.98%, and the model's mAP was 92.19%. The average detection time of a single image was 35.33ms. (2) The YOLO-Banana-l4 model reduced the weight and the detection time compared to the YOLOv4 model, but it was not selected finally because the detection accuracy was lower than that of the YOLO-Banana model. (3) In banana orchard detection, the texture and the size of banana bunches is more obvious than that of the stalk; for this reason, the correct detection rate of banana bunches is greater than that of the stalks. Small-sized bunches and stalks, the degree of occlusion, and the banana variety are several details worth considering in the detection task. The improved model proposed in this study is robust to illumination and shows satisfactory detection performance in different occlusion environments. In the future work, the banana bunches' and stalks' locations will be realized on the basis of the detection model, and the real coordinates will be obtained to provide information for the management of the banana orchard. Furthermore, the detection of different species of bananas will be conducted by collecting more banana images.

Author Contributions: Conceptualization, Z.Y. and X.Z.; methodology, L.F.; software, L.F.; validation, F.W. and J.L.; formal analysis, L.F.; investigation, Y.C.; resources, J.D.; data curation, L.F.; writing—original draft preparation, L.F.; writing—review and editing, X.Z.; visualization, F.W. and J.L.; supervision, X.Z and Z.Y.; project administration, J.D.; funding acquisition, J.D. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2020YFD1000104), China Agriculture Research System of MOF and MARA (CARS31-10), and the Guangzhou Key Research and Development Program (No. 107208411092).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to Lei Zhou of the South China Robotics Innovation Research Institute and the laboratory staff Yunhe Ding, Yinlong Jiang, Lei Zhao, and Lizhe Ma for their contributions, constructive suggestions, and valuable help.

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

References

1. Stevens, B.; Diels, J.; Brown, A.; Bayo, S.; Ndakidemi, P.A.; Swennen, R. Banana Biomass Estimation and Yield Forecasting from Non-Destructive Measurements for Two Contrasting Cultivars and Water Regimes. *Agronomy* **2020**, *10*, 1435. [CrossRef]
2. Gongal, A.; Amatya, S.; Karkee, M.; Zhang, Q.; Lewis, K. Sensors and systems for fruit detection and localization: A review. *Comput. Electron. Agric.* **2015**, *116*, 8–19. [CrossRef]
3. Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Li, J.; Lian, G.; Zou, X. Recognition and localization methods for vision-based fruit picking robots: A review. *Front. Plant Sci.* **2020**, *11*, 510. [CrossRef]
4. Wang, C.; Tang, Y.; Zou, X.; Luo, L.; Chen, X. Recognition and Matching of Clustered Mature Litchi Fruits Using Binocular Charge-Coupled Device (CCD) Color Cameras. *Sensors* **2017**, *17*, 2564. [CrossRef]
5. Fu, L.; Tola, E.; Al-Mallahi, A.; Li, R.; Cui, Y. A novel image processing algorithm to separate linearly clustered kiwifruits. *Biosyst. Eng.* **2019**, *183*, 184–195. [CrossRef]
6. Reis, M.J.C.S.; Morais, R.; Peres, E.; Pereira, C.; Contente, O.; Soares, S.; Valente, A.; Baptista, J.; Ferreira, P.J.S.G.; Bulas Cruz, J. Automatic detection of bunches of grapes in natural environment from color images. *J. Appl. Log.* **2012**, *10*, 285–290. [CrossRef]
7. Cubero, S.; Diago, M.P.; Blasco, J.; Tardáguila, J.; Millán, B.; Aleixos, N. A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis. *Biosyst. Eng.* **2014**, *117*, 62–72. [CrossRef]
8. Wang, C.; Lee, W.S.; Zou, X.; Choi, D.; Gan, H.; Diamond, J. Detection and counting of immature green citrus fruit based on the Local Binary Patterns (LBP) feature using illumination-normalized images. *Precis. Agric.* **2018**, *19*, 1062–1083. [CrossRef]
9. Nuske, S.; Wilshusen, K.; Achar, S.; Yoder, L.; Narasimhan, S.; Singh, S. Automated Visual Yield Estimation in Vineyards. *J. Field Robot.* **2014**, *31*, 837–860. [CrossRef]
10. Yamamoto, K.; Guo, W.; Yoshioka, Y.; Ninomiya, S. On Plant Detection of Intact Tomato Fruits Using Image Analysis and Machine Learning Methods. *Sensors* **2014**, *14*, 12191–12206. [CrossRef]
11. Tao, Y.; Zhou, J. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. *Comput. Electron. Agric.* **2017**, *142*, 388–396. [CrossRef]
12. Fu, L.; Duan, J.; Zou, X.; Lin, G.; Song, S.; Ji, B.; Yang, Z. Banana detection based on color and texture features in the natural environment. *Comput. Electron. Agric.* **2019**, *167*, 105057. [CrossRef]
13. Zhao, Y.; Gong, L.; Huang, Y.; Liu, C. A review of key techniques of vision-based control for harvesting robot. *Comput. Electron. Agric.* **2016**, *127*, 311–323. [CrossRef]
14. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
15. Alex, K.; Ilya, S.; Geoffrey, E.H. ImageNet Classification with Deep Convolutional Neural Networks. In *NIPSNIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012*; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 25, pp. 1097–1105.
16. Altaheri, H.; Alsulaiman, M.; Muhammad, G. Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning. *IEEE Access* **2019**, *7*, 117115–117133. [CrossRef]
17. Liu, Z.; Wu, J.; Fu, L.; Majeed, Y.; Feng, Y.; Li, R.; Cui, Y. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. *IEEE Access* **2020**, *8*, 2327–2336. [CrossRef]
18. Kang, H.; Chen, C. Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards. *Sensors* **2019**, *19*, 4599. [CrossRef]
19. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]
20. Wang, Y.; Lv, J.; Xu, L.; Gu, Y.; Zou, L.; Ma, Z. A segmentation method for waxberry image under orchard environment. *Sci. Hortic.* **2020**, *266*, 109309. [CrossRef]
21. Chen, T.; Zhang, R.; Zhu, L.; Zhang, S.; Li, X. A method of fast segmentation for banana stalk exploited lightweight multi-feature fusion deep neural network. *Machines* **2021**, *9*, 66. [CrossRef]
22. Majeed, Y.; Karkee, M.; Zhang, Q.; Fu, L.; Whiting, M.D. Determining grapevine cordon shape for automated green shoot thinning using semantic segmentation-based deep learning networks. *Comput. Electron. Agric.* **2020**, *171*, 105308. [CrossRef]
23. Li, J.; Tang, Y.; Zou, X.; Lin, G.; Wang, H. Detection of Fruit-bearing Branches and Localization of Litchi Clusters for Vision-based Harvesting Robots. *IEEE Access* **2020**, *8*, 117746–117758. [CrossRef]
24. Chen, M.; Tang, Y.; Zou, X.; Huang, K.; Huang, Z.; Zhou, H.; Wang, C.; Lian, G. Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology. *Comput. Electron. Agric.* **2020**, *174*, 105508. [CrossRef]
25. Lin, G.; Tang, Y.; Zou, X.; Xiong, J.; Li, J. Guava detection and pose estimation using a low-cost RGB-D sensor in the field. *Sensors* **2019**, *19*, 428. [CrossRef]
26. Mu, Y.; Chen, T.; Ninomiya, S.; Guo, W. Intact detection of highly occluded immature tomatoes on plants using deep learning techniques. *Sensors* **2020**, *20*, 2984. [CrossRef]
27. Neupane, B.; Horanont, T.; Hung, N.D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLoS ONE* **2019**, *14*, e0223906. [CrossRef]
28. Cheng, Z.; Zhang, F. Flower End-to-End Detection Based on YOLOv4 Using a Mobile Device. *Wirel. Commun. Mob. Comput.* **2020**, *2020*, 1–9. [CrossRef]

29. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors* **2016**, *16*, 1222. [CrossRef]
30. Vasconez, J.P.; Delpiano, J.; Vougioukas, S.; Cheein, F.A. Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. *Comput. Electron. Agric.* **2020**, *173*, 105348. [CrossRef]
31. Chen, M.; Tang, Y.; Zou, X.; Huang, Z.; Zhou, H.; Chen, S. 3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM. *Comput. Electron. Agric.* **2021**, *187*, 106237. [CrossRef]
32. Boogaard, F.P.; Rongen, K.S.A.H.; Kootstra, G.W. Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging. *Biosyst. Eng.* **2020**, *192*, 117–132. [CrossRef]
33. Suo, R.; Gao, F.; Zhou, Z.; Fu, L.; Song, Z.; Dhupia, J.; Li, R.; Cui, Y. Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking. *Comput. Electron. Agric.* **2021**, *182*, 106052. [CrossRef]
34. Xie, H.; Dai, N.; Yang, X.; Zhan, K.; Liu, J. Research on recognition methods of pomelo fruit hanging on trees base on machine vision. In *2019 ASABE Annual International Meeting*; American Society of Agricultural and Biological Engineers: Boston, MA, USA, 2019; p. 1900411.
35. Santos, T.T.; de Souza, L.L.; Dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron. Agric.* **2020**, *170*, 105247. [CrossRef]
36. Fu, L.; Duan, J.; Zou, X.; Lin, J.; Zhao, L.; Li, J.; Yang, Z. Fast and accurate detection of banana fruits in complex background orchards. *IEEE Access* **2020**, *8*, 196835–196846. [CrossRef]
37. Zhang, R.; Li, X.; Zhu, L.; Zhong, M.; Gao, Y. Target detection of banana string and fruit stalk based on YOLOv3 deep learning network. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE2021), Nanchang, China, 26–28 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 346–349.
38. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precis. Agric.* **2019**, *20*, 1107–1135. [CrossRef]
39. Liu, G.; Nouaze, J.C.; Touko Mbouembe, P.L.; Kim, J.H. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. *Sensors* **2020**, *20*, 2145. [CrossRef]
40. Lawal, O.M. YOLOMuskmelon: Quest for Fruit Detection Speed and Accuracy Using Deep Learning. *IEEE Access* **2021**, *9*, 15221–15227. [CrossRef]
41. Wu, F.; Duan, J.; Chen, S.; Ye, Y.; Ai, P.; Yang, Z. Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point. *Front. Plant Sci.* **2021**, *12*, 705021. [CrossRef]
42. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [CrossRef]