agriculture

MDPI

Article

Instance Segmentation of Lotus Pods and Stalks in Unstructured Planting Environment Based on Improved YOLOv5

Ange Lu [1,2], Lingzhi Ma [1,2], Hao Cui [1,2,*], Jun Liu [1,2] and Qiucheng Ma [1,2]

[1] School of Mechanical Engineering and Mechanics, Xiangtan University, Xiangtan 411105, China; ange.lu@xtu.edu.cn (A.L.); 202121541937@smail.xtu.edu.cn (L.M.); 202121542019@smail.xtu.edu.cn (J.L.); mqc@xtu.edu.cn (Q.M.)

[2] Engineering Research Center of Complex Track Processing Technology & Equipment, Ministry of Education, Xiangtan University, Xiangtan 411105, China

* Correspondence: 202221542098@smail.xtu.edu.cn

Abstract: Accurate segmentation of lotus pods and stalks with pose variability is a prerequisite for realizing the robotic harvesting of lotus pods. However, the complex growth environment of lotus pods causes great difficulties in conducting the above task. In this study, an instance segmentation model, LPSS-YOLOv5, for lotus pods and stalks based on the latest YOLOv5 v7.0 instance segmentation model was proposed. The CBAM attention mechanism was integrated into the network to improve the model's feature extraction ability. The scale distribution of the multi-scale feature layer was adjusted, a 160 × 160 small-scale detection layer was added, and the original 20 × 20 large-scale detection layer was removed, which improved the model's segmentation accuracy for small-scale lotus stalks and reduced the model size. On the medium-large scale test set, LPSS-YOLOv5 achieved a mask $mAP_{0.5}$ of 99.3% for all classes. On the small-scale test set, the $mAP_{0.5}$ for all classes and $AP_{0.5}$ for stalks were 88.8% and 83.3%, which were 2.6% and 5.0% higher than the baseline, respectively. Compared with the mainstream Mask R-CNN and YOLACT models, LPSS-YOLOv5 showed a much higher segmentation accuracy, speed, and smaller size. The 2D and 3D localization tests verified that LPSS-YOLOv5 could effectively support the picking point localization and the pod–stalk affiliation confirmation.

Keywords: lotus pods; instance segmentation; deep learning; YOLOv5; attention mechanism

Citation: Lu, A.; Ma, L.; Cui, H.; Liu, J.; Ma, Q. Instance Segmentation of Lotus Pods and Stalks in Unstructured Planting Environment Based on Improved YOLOv5. *Agriculture* **2023**, *13*, 1568. https://doi.org/10.3390/agriculture13081568

Academic Editor: Xanthoula Eirini Pantazi

Received: 20 July 2023
Revised: 31 July 2023
Accepted: 2 August 2023
Published: 6 August 2023

1. Introduction

Lotus seeds are mature seeds of the aquatic herbaceous economic plant Lotus, distributes mainly in Asian, Australian, and North American countries. Due to their delicious taste, rich nutrients, and medicinal ingredients, they are used as fresh-eating fruit, functional food, and traditional Chinese medicine [1–4]. The lotus seeds grow inside the lotus pod, which is bowl-shaped and supported by a slender lotus stalk (Figure 1).

Lotus pod harvesting is an essential procedure of the lotus seed cultivation process. Limited to the aquatic or muddy growth environment and gradual ripening and random location distribution characteristics of the lotus pods, manual selective picking has been the only way to harvest lotus pods since ancient times. However, manual labor involves harsh working conditions, high labor intensity, and low efficiency. Especially, the mature period of lotus seeds is in the high-temperature season between July and September every year [5]. Many experienced workers are often required to participate in picking during this period. With the intensification of urbanization and population aging, the labor shortage for harvesting has become prominent, and the harvesting cost has subsequently increased [6,7]. Hence, automatic lotus pod-harvesting technology is urgently needed to fulfill the requirement of the lotus seeds industry.

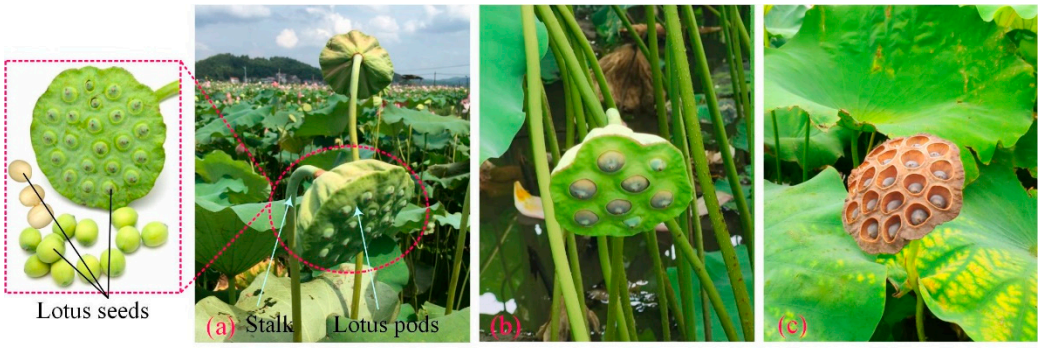

Figure 1. Appearance of the mature lotus pods in the natural growth environment. (**a**) Green ripening stage. (**b**) Full ripening stage. (**c**) Overripe stage.

At present, robotic picking technology has become a popular research field in the automatic harvesting of fruits and vegetables, which provides an effective solution to the challenges faced in manual picking [8,9]. For manual lotus pod picking, the lotus pod is separated from the lotus stalk by manually breaking or cutting off the stalk near the lotus pod. Therefore, to realize effective robotic lotus pod picking, it is necessary to identify and segment the lotus stalk area to provide basic data for the end-effector to perform the holding and separation actions. On the other hand, to prevent misjudgment due to the similarity between the characteristics of the lotus stalk and the surrounding lotus leaf stalk and the possible gathering of multiple lotus pods, affiliation analysis of the lotus pod and corresponding stalk is needed in combination with lotus pod identification information before automatic picking. Hence, studying fast, robust, and effective segmentation methods for lotus pods and stalks is necessary.

With the development of harvesting robot technology, image segmentation of fruits and vegetables has received extensive attention from researchers, with relevant literature reports divided into traditional image processing and deep learning. The traditional image processing-based segmentation methods preprocess the image first and then utilize the differences in color, shape, and texture between the object and the background in the image to achieve segmentation [10–15]. Septiarini et al. [11] proposed an image segmentation method for oil palm fruit, which includes procedures of object localization, color and smoothing pre-processing, and edge detection. Linker et al. [12] developed a four-step detection and segmentation method for green apples based on color, texture, and contour information.

Although the segmentation method based on traditional image processing is simple and convenient, it relies on high-quality image processing and hand-engineering features and is sensitive to environmental changes. Typically, there are problems in the growth environment of agricultural products, such as illumination changes, occlusion of branches and leaves, overlapping, and similarity in color between objects and background, which have posed significant challenges to the traditional method.

In recent years, deep learning technology with high precision, high efficiency, and good robustness has been applied in the segmentation of fruits and vegetables [10,16]. It can realize autonomous learning, automatic extraction of image feature information, and end-to-end detection [17,18] and is more suitable for segmentation tasks in complex environments than the traditional method. The existing segmentation method includes three sub-types: semantic segmentation [19], instance segmentation, and panoptic segmentation [20], among which instance segmentation could obtain pixel-level masks of individual objects [21] and is the primary type for object segmentation in automatic harvesting applications.

The instance segmentation algorithms reported mainly include two-stage algorithms represented by Mask R-CNN [22,23] and one-stage algorithms represented by You Only Look at CoefficienTs (YOLACT) [24], UNet, fully convolutional one-stage (FCOS), etc. For the segmentation of fruits and leaves, Wang and He [10] designed an apple segmentation method by integrating the attention mechanism module into the backbone network of the

Mask R-CNN algorithm. A segmentation mAP of 0.917 was achieved. Lu et al. [14] studied the application of deep learning instance segmentation algorithms, i.e., the pyramid scene parsing network (PSPNet), U-Net, and DeepLabV3+, on the segmentation task of Sichuan peppers. They also compared the results with three traditional segmentation methods, i.e., RGB and HSV color spaces, and k-means clustering. Xu et al. [25] segmented cherry tomatoes and stems using an improved Mask R-CNN algorithm, achieving identification accuracies of 93.76% and 89.34%, respectively. Jia et al. [26] proposed a segmentation method for green fruits called FoveaMask, which introduced a position attention module and performed instance segmentation of fruit through the full convolutional operation. Liu et al. [27] proposed a modified FCOS model for the segmentation of obscured green fruit and achieved a segmentation accuracy of 85.3% on an Apple dataset.

For the segmentation of fruit peduncles and branches, Li et al. [28] proposed a multitask-aware YOLACT network, which realized the segmentation of the main stem and fruit pedicel of cherry tomatoes. Zhong et al. [29] studied the segmentation of the main fruit bearing branches (MFBB) of litchi using YOLACT. Yang et al. [30] developed an integrated system that simultaneously detects and measures citrus fruits and branches by combining Mask R-CNN and a branch segmentation fusion algorithm. The achieved average accuracy of fruit and branch identification was 88.15% and 96.27%, respectively. Hitherto, to the authors' knowledge, there are very few reports on the segmentation of lotus pods and stalks.

The segmentation task of lotus pods and stalks in an unstructured planting environment is quite difficult due to color similarity between the objects and background, scattered distribution and multi-scale characteristics, and occlusion phenomenon. In this study, the effective instance segmentation of lotus pods and stalks in the unstructured planting environment is investigated. A specific segmentation dataset of lotus pods and stalks is established. A model, LPSS-YOLOv5 (Lotus Pod and Stalk Segmentation Model Based on You Only Look Once version 5), for lotus pod and stalk segmentation based on the latest YOLOv5 v7.0 instance segmentation model is proposed. The CBAM (Convolutional Block Attention Module) attention mechanism is introduced into the model network. And the scale distribution of the network output feature layer is adjusted. Then, the model is compared with the mainstream Mask R-CNN and YOLACT segmentation models. In addition, a method for localizing the stalk picking point and lotus pod's key point in the 2D image based on the model's segmentation result is built and tested. A 3D localization test is conducted based on the self-developed lotus pod harvesting robot. The research results are expected to support the development of the lotus pod harvesting robots.

2. Materials and Methods

2.1. Image Collection

At present, there is still a lack of dedicated segmentation datasets for lotus pods and stalks that could be directly utilized. For this purpose, image collection work of lotus pods in actual planting environments was carried out in this study. The image collection locations were located in several lotus seed planting bases in Xiangtan City, Hunan Province, China, and the collection time was July and August 2021, and July 2022. The varieties of lotus seeds used were '*TaiKong*' and '*CunSan*', and the collection objects were mature lotus pods ready for harvest (Figure 1). The climatic conditions during the collection included sunny, rainy, and overcast; the time periods were morning, forenoon, and afternoon. Various types of smartphones with cameras were used as acquisition devices. During the shooting process, it was ensured that the lotus pods and stalks in the images were clear, and different lighting conditions, such as direct sunlight, backlight, shadow, and cloudy, were considered. Some examples of the lotus pod images under different lighting conditions is shown in Figure 2.

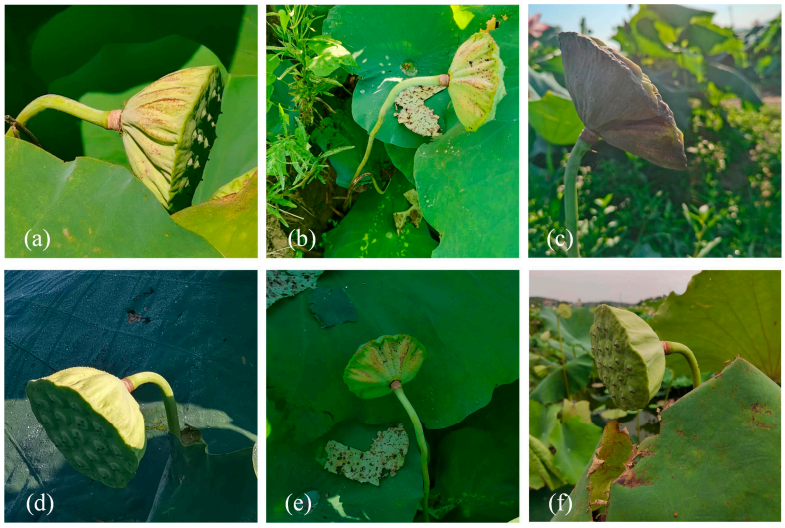

Figure 2. Examples of the lotus pod images under different lighting conditions. (**a**,**b**) sunny and direct sunlight environments. (**c**) Sunny and backlight environments. (**d**,**e**) Partially or completely in the shadow of surrounding objects. (**f**) Cloudy environment.

2.2. Dataset Preparation

A total of 2500 images were prepared as the dataset for model training, which was then divided into a training set (2250 images) and a validation set (250 images) at a ratio of 9:1. Due to the randomness of the size and location distribution of lotus pod individuals in the growth environment, multi-scale features are reflected in the images. Therefore, a combined test set was built independently in this study to comprehensively test the segmentation performance of the deep learning instance segmentation models, wherein 380 images containing medium-large scale lotus pod objects were prepared as the medium-large scale test set (Test Set A). In addition, 320 images of mostly long-distance or small-sized lotus pods were used to make the small-scale test set (Test Set B). The ratio of the lotus pod's mask area over the whole image area was used to discriminate the scales. The utilized thresholds between medium and large, medium and small scales were 0.125 and 0.025, respectively. Figure 3 shows some lotus pods with different scales.

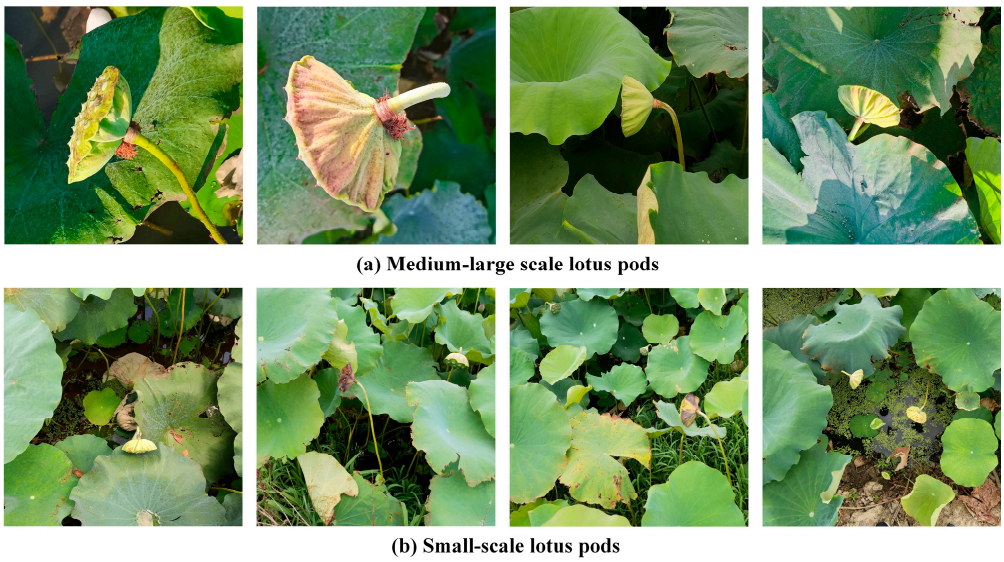

Figure 3. Illustration of the lotus pods with different scales.

In order to enhance the diversity of the datasets and the robustness of the model, the number of images in the training set was expanded to 4500 using data augmentation.

Four augmentation methods were used, including adding salt-and-pepper noise, adding Gaussian noise, changing brightness, and flipping the image. The specific operation method was to randomly implement one to four augmentation methods on a single image and ensure that at least one augmentation effect was arranged for each image. Figure 4 illustrates the augmentation methods adopted and the corresponding augmentation effects.

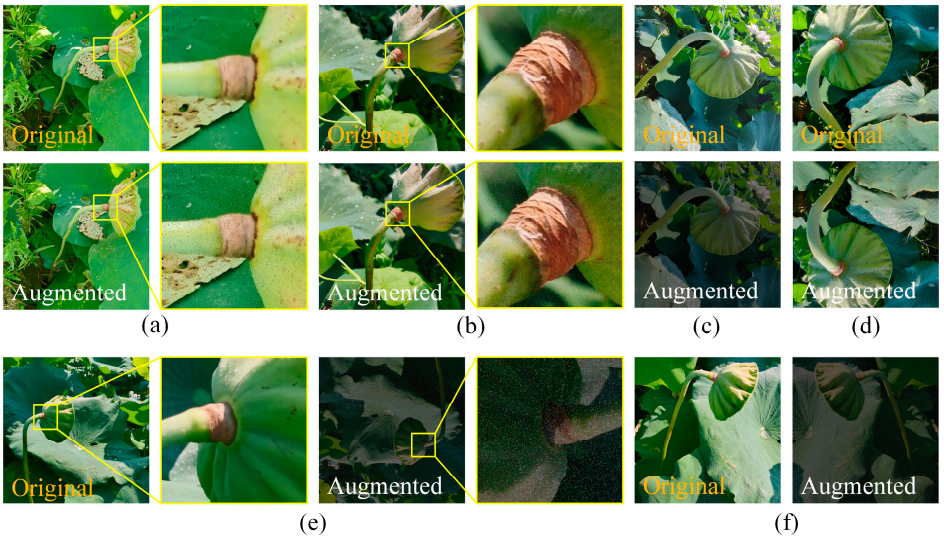

Figure 4. Various data augmentation methods adopted. (**a**) Adding salt-and-pepper noise. (**b**) Adding Gaussian noise. (**c**) Changing brightness. (**d**) Flipping the image. (**e**,**f**) Multiple augmentation methods superimposed on one image.

LabelMe annotation software was used to label the contours of the lotus pods and the corresponding lotus stalks in the dataset, with the category labels for lotus pods and stalks being labeled as 'LP' and 'Stalk', respectively. The labeling principle was to annotate the lotus pods and stalks that could be manually distinguished in the image. For lotus pods, the polygon labeling area enveloped the entire lotus pod entity contour; for stalks, the polygon labeling area started from the lotus pod pedicel and ended at the bending part of the stalk, as shown in Figure 5. Additionally, for cases where only the lotus pod could be observed, only the lotus pod in the image was annotated (Figure 5b). Table 1 shows the statistics of the number of images and labels corresponding to each dataset.

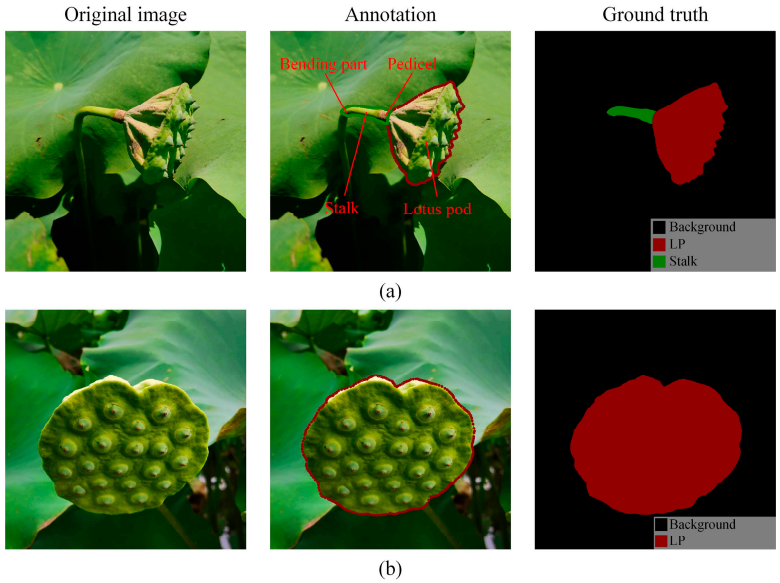

Figure 5. Sample annotation diagrams of (**a**) lotus pod and stalk; (**b**) lotus pod only.

Table 1. Label statistics of lotus pods and stalks of the datasets.

Dataset	Image Count	Label Count		
		LP	Stalk	All
Training	4500	6612	5408	12,020
Validation	250	370	316	686
Test Set A	380	422	380	802
Test Set B	320	864	668	1532

2.3. Instance Segmentation Method of Lotus Pods and Stalks Based on YOLOv5

2.3.1. Overview of YOLOv5 v7.0 Instance Segmentation Model

YOLOv5 v7.0 is the state-of-the-art real-time instance segmentation algorithm [31]. Its network consists of the input end, backbone network, neck network, and head network parts, as shown in Figure 6. The input end is used to receive images and perform preprocessing operations on the images, including mosaic data enhancement, adaptive image scaling, and adaptive anchor box calculation.

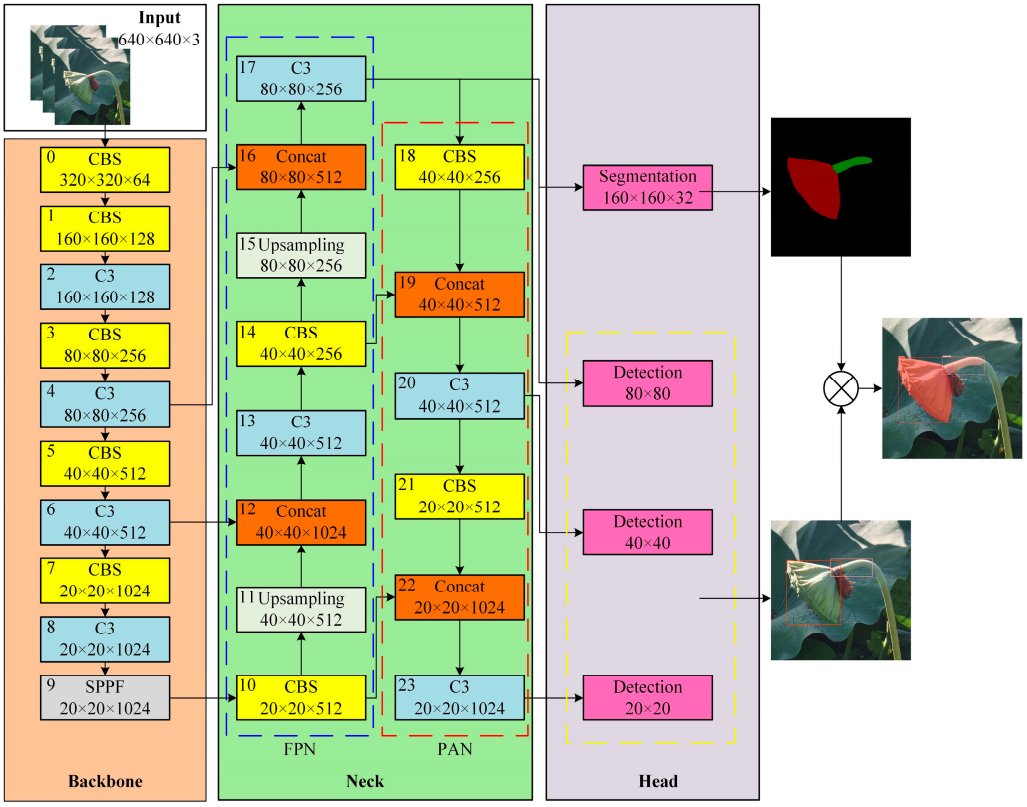

Figure 6. Network structure of YOLOv5 v7.0 (model s) instance segmentation model.

The backbone network is used to extract features from the image, consisting of multiple CBS (convolution, BN layer, and SiLU activation function), C3 (concentrated-comprehensive convolution), and SPPF (Spatial Pyramid Pooling-Fast) layers.

The neck network combines the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN) to achieve multi-scale feature fusion of the features extracted through the backbone network to obtain rich feature information.

The head network is divided into two branches: the object detection head and the instance segmentation head, wherein, the object detection head inherits the multi-scale feature fusion mechanism of the ordinary YOLOv5 detection model (e.g., version 6.0) and realizes the detection of objects with different sizes at three feature scales of 20 × 20, 40 × 40, and 80 × 80. The instance segmentation head achieves pixel-by-pixel classification

prediction and generates binary masks for the objects through a small fully convolutional neural network (FCN). The segmentation head is a 4-layer network structure composed of CBS and up-sampling modules, and the structure is shown in Figure 7a. It uses the output of the C3 module of the 17th layer of the neck network as the input. First, the first CBS module (with a 3 × 3 convolution kernel and a stride of 1) processes the input feature map and outputs a feature map with a size of 80 × 80 × 256. This feature map is then up-sampled once to expand the size to 160 × 160 × 256. Subsequently, it passes through two CBS modules in succession (the convolution kernels are 3 × 3 and 1 × 1, respectively, and the strides are all 1), and finally, its size is reduced to 160 × 160 × 32 and outputted. The above network structure enables the YOLOv5 v7.0 model to achieve high segmentation accuracy and efficiency.

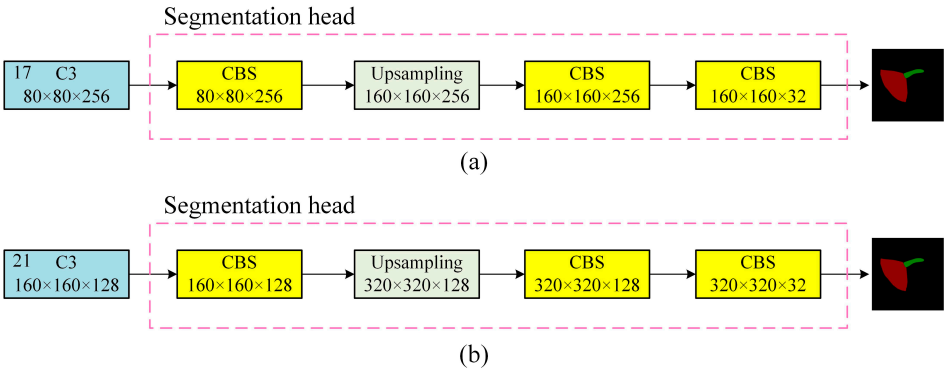

Figure 7. Network structure of the instance segmentation head. (**a**) YOLOv5 v7.0. (**b**) LPSS-YOLOv5.

2.3.2. Improvements in the Proposed LPSS-YOLOv5 Instance Segmentation Model

The growth environment of lotus pods is highly unstructured. Compared with those standardized planted fruits, e.g., strawberry, apple, cherry tomato, pepper, and kiwifruit, the difficulty in object detection and segmentation is dramatically increased. First, the color of the lotus pod is similar to the surrounding objects, e.g., lotus leaves, weeds, and soil, and it is easily occluded. Second, lotus pods are characterized by gradual ripening and scattered distribution in the planting environment. Consequently, there are differences in size, shape, and maturity among different individuals within the same area. Third, there are spatial overlapping phenomena between lotus pods and lotus pods, lotus pods and lotus leaves. Due to the existence of the above factors, even manual on-site identification presents significant challenges. On the other hand, many mainstream instance segmentation models, including YOLOv5 v7.0, were designed and tested based on the standard COCO dataset. However, the object types in the COCO dataset do not include the lotus pod. The difference between the lotus pod and the objects in the COCO dataset will affect the segmentation effect of the YOLOv5 v7.0 model for lotus pods and stalks. Therefore, it is necessary to develop a specialized instance segmentation model for lotus pod and stalk segmentation.

In this study, an instance segmentation model, LPSS-YOLOv5, was proposed for lotus pod and lotus stalk objects in actual planting environments. The overall network structure of the model is shown in Figure 8. Compared to the original YOLOv5 v7.0 model, two main improvements were carried out: (a) introducing the CBAM attention mechanism into the model network; and (b) adjusting the scale distribution of the multi-scale feature layer.

(a) Introduction of the CBAM attention mechanism

It is difficult to detect and segment lotus pods and corresponding stalks in actual planting environments, especially for small-scale individuals. The attention mechanism is an improvement method in deep learning mainly used to enhance the model's feature extraction ability for objects in complex environments and thus improve detection performance. CBAM is a lightweight, general-purpose attention mechanism for the feed-forward convolutional neural network. It strengthens the extraction of important information in the feature map by combining channel and spatial attention, and the ability to suppress

irrelevant information. The CBAM module could be conveniently integrated into the modules of existing convolutional neural networks and has been widely applied in various object detection tasks [32,33].

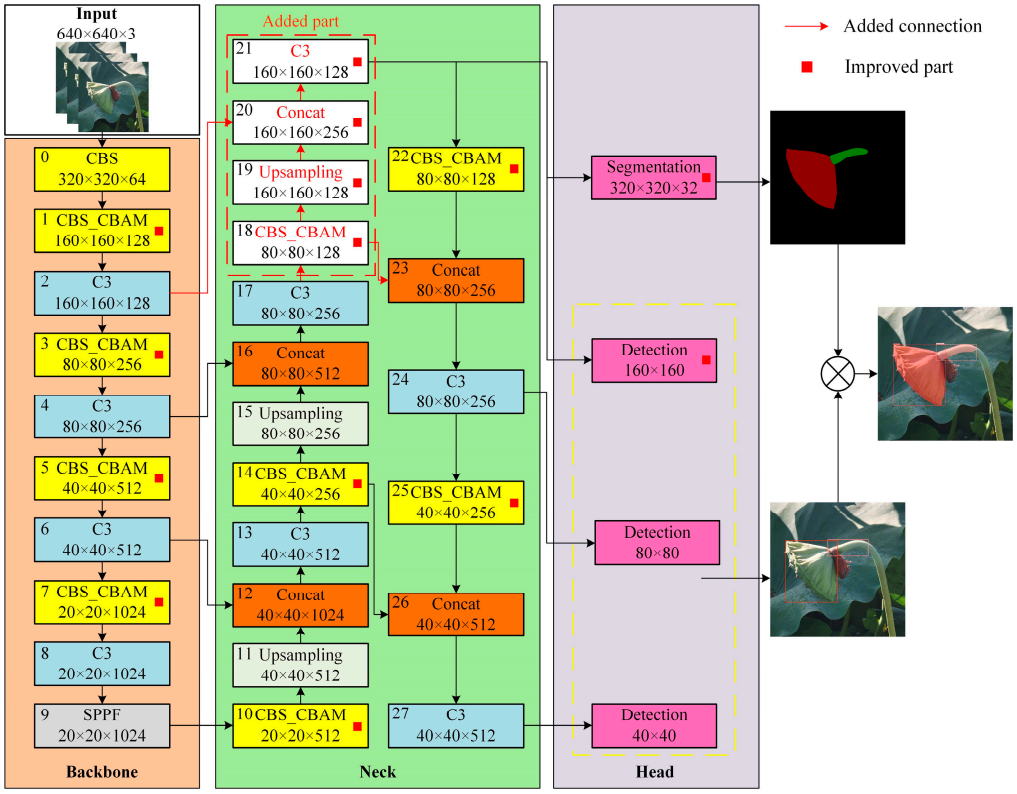

Figure 8. Network structure of the LPSS-YOLOv5 instance segmentation model.

CBAM consists of two sub-modules: channel attention and spatial attention, as shown in Figure 9. For each intermediate feature map F inputted into the CBAM module, it is sequentially passed through the channel and spatial attention modules to generate a refined feature map output, where the channel attention focuses on 'what' the useful information is, while spatial attention focuses on 'where' the useful information is.

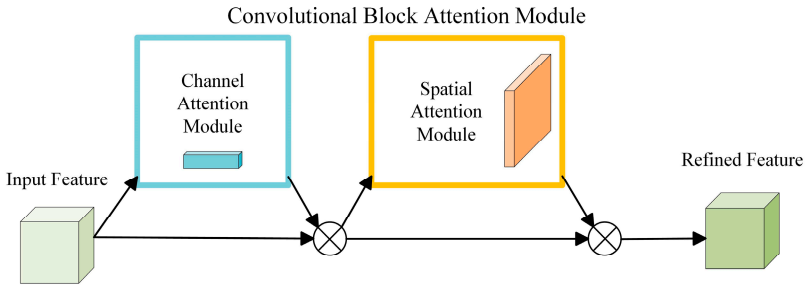

Figure 9. Structure of the CBAM.

In this study, the CBAM attention mechanism was introduced into the LPSS-YOLOv5 network to enhance the model's sensitivity and feature extraction ability for lotus pods and stalks. In the network, except for the CBS module of layer 0 in the original YOLOv5 network, all other CBS modules were replaced by CBS_CBAM modules integrated with the CBAM, as shown in Figure 8. The structure of the CBS_CBAM module is shown in Figure 10, which consists of the Conv2d module (with convolution kernel of 1×1, stride of 1), batch normalization (BN) layer, Sigmoid weighted linear unit activation function (SiLU), and the CBAM module.

Figure 10. Structure of the CBS_CBAM module.

(b) Adjustment of the scale distribution of the multi-scale feature layer

The original YOLOv5 uses three feature map sizes of 20 × 20, 40 × 40, and 80 × 80 to achieve multi-scale detection of large, medium, and small objects, respectively. However, even with adopting the 80 × 80 small-scale detection layer, the original YOLOv5 model still may fail to effectively detect some small-scale lotus pods and stalks in planting environments, thus affecting the overall detection rate of the harvesting robot. To improve this situation, we adjusted the scale distribution of the multi-scale feature layers of the LPSS-YOLOv5 network. A small object detection layer of 160 × 160 was added to improve the model's segmentation accuracy for small-scale lotus pod and stalk objects. Moreover, the original 20 × 20 large-scale detection layer was removed to reduce the model's parameter amount and size.

Specifically, as shown in Figure 8, first, a CBS_CBAM layer, an up-sampling layer, a Concat layer, and a C3 module were added after the 17th layer's C3 module in the original YOLOv5 network. The CBS_CBAM layer outputted the feature map of the 18th layer to the Concat in the 23rd layer and the up-sampling layer in the 19th layer. Next, the up-sampling layer conveyed the feature map to the Concat layer in the 20th layer. The Concat layer concatenated the output feature map of the C3 module in the 2nd layer with the output feature map of the 19th layer. The feature map was conveyed to the instance segmentation head, detection head, and the 22nd layer, respectively, after the 21st layer's C3 module processing. After the above calculation, the network ultimately formed a small object detection head with a scale of 160 × 160 and an improved segmentation head whose input and output feature map sizes were 160 × 160 × 128 and 320 × 320 × 32, respectively (Figure 7b).

It is worth noting that after the network adjusted the scale distribution of the multi-scale feature layer, it better adapted to the characteristics of small-scale lotus pods and stalks that are sparsely distributed and occupy a small proportion in the image. By adding the CBS_CBAM modules, the network could pay adequate attention to important information in local regions. At the same time, the up-sampling and Concat operations enable the effective fusion of shallow and deep features so that the feature map could better capture the detailed information of small objects, thereby improving the detection and segmentation performance of the network on the small scale and overlapping objects.

2.4. Experimental Preparation

All model training and experiments in this study were run on a computer configured with Intel Core i7-13700K CPU, GeForce RTX 4090 GPU, and 32GB of memory. Regarding software, the operating system was Windows 11, the CUDA version was 11.6, the deep learning framework was PyTorch 1.12, and the programming language used was Python 3.7. The basic model training parameters are listed in Table 2.

Table 2. Model training parameters of the used instance segmentation models.

Parameters	Instance Segmentation Models			
	LPSS-YOLOv5	**YOLOv5 v7.0**	**Mask R-CNN**	**YOLACT**
Backbone	CSPDarknet53	CSPDarknet53	ResNet50	ResNet101
Input size	640 × 640	640 × 640	640 × 640	640 × 640
Learning rate	0.01	0.01	0.02	0.01
Batch size	16	16	64	25
Epoch/Iteration	200 (Epoch)	200 (Epoch)	60 (Epoch)	8000 (Iter.)
Momentum coefficient	0.9	0.9	0.9	0.9

2.5. Evaluation Metrics of the Model

The following evaluation metrics of precision (P), recall (R), F1-score, mean Average Precision (mAP), Frames Per Second (FPS), and Mean Intersection over Union (mIOU) were used to validate the segmentation performance of the above models. Among them, the mAP is the average AP value for all object categories to characterize the comprehensive instance segmentation accuracy of the model. And the IOU threshold of the mAP indicator was set to 0.5 (i.e., $mAP_{0.5}$) in this study. The F1-score is the harmonic mean of P and R. The FPS represents the real-time performance of the model. The mIOU is the average of the IOU of the predicted mask pixel area and the ground truth for all categories, reflecting the coverage quality of the mask. The calculation methods of the above indicators are as follows. When the values were higher, the performance of the model was better.

$$P = \frac{TP}{TP + FP} \times 100\% \tag{1}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

$$F1 - score = \frac{2 \times P \times R}{(P + R)} \tag{3}$$

$$AP_i = \int_0^1 P(R)dR \tag{4}$$

$$mAP_i = \frac{1}{k}\sum_{i=1}^{k} AP_i \tag{5}$$

where TP is the number of true positive masks; FP is the number of false positive masks; FN is the number of false negative masks; and k is the number of labeled categories.

$$mIOU = \frac{1}{k}\sum_{i=1}^{k} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{6}$$

where TP_i is the total number of pixels classified and labeled as category i. FN_i is the number of pixels, which are labeled as category i, but classified as other categories. FP_i is the number of pixels, which are labeled as other categories, but classified as class i.

3. Results and Discussion

3.1. Instance Segmentation Performance of the Proposed LPSS-YOLOv5

In this section, experiments were carried out to test the performance of the proposed LPSS-YOLOv5 model. The model's loss curves on the training and validation sets are shown in Figure 11. In line with the original YOLOv5 7.0 model, the loss curve combined four loss components, namely box loss, segmentation loss, object loss, and classification loss. In Figure 11, both the training and validation loss curves were ideal L-shaped. With the increase in training times, the loss value of each curve gradually decreases until final convergence, which indicates that the model's performance has stabilized and the whole training process was normal. In addition, the gap between the training and validation loss curves was small, and the model fitted well.

Then, the model was tested using Test Sets A and B to verify its segmentation performance for lotus pods and stalks in actual planting environments. The trained weight was deployed to the model, and then the images in the test sets were inputted into the model for inference, and test results were obtained accordingly, as listed in Table 3.

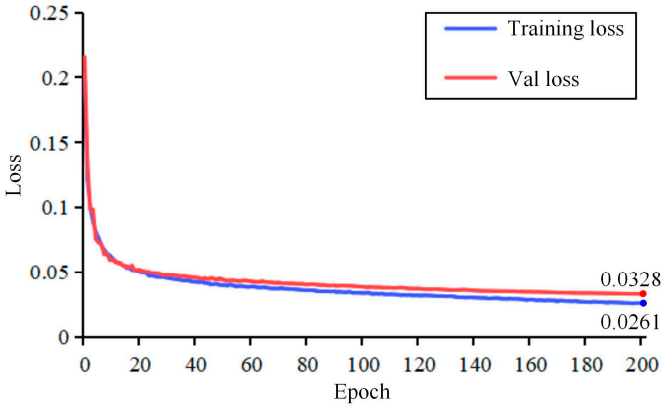

Figure 11. Loss curves of the LPSS-YOLOv5 model.

Table 3. Segmentation performance of the proposed LPSS-YOLOv5 model.

Evaluation Metrics (Mask)	Test Set A			Test Set B		
Label Class	All	LP	Stalk	All	LP	Stalk
P (%)	96.7	95.6	97.9	93.4	95.4	91.3
R (%)	99.3	99.8	98.8	83.0	88.7	77.4
F1-score	98.0	97.7	98.3	87.9	91.9	83.8
$AP_{0.5}$ or $mAP_{0.5}$ (%)	99.3	99.3	99.3	88.8	94.2	83.3

The P, R, F1-score, and $mAP_{0.5}$ values achieved with LPSS-YOLOv5 on Test Set A were 96.7%, 99.3%, 98.0%, and 99.3%, respectively. Among them, the $AP_{0.5}$ of the model for the lotus pod and stalk categories were both 99.3%. This means that the segmentation performance of the model for medium-large scale lotus pods and stalks is high. On Test Set B, the P, R, F1-score, and $mAP_{0.5}$ values achieved with the model were 93.4%, 83.0%, 87.9%, and 88.8%, respectively. Among them, the $AP_{0.5}$ corresponding to the lotus pod and stalk categories were 94.2% and 83.3%, respectively.

Further, on Test Set B, the segmentation performance of the model was lower than those of Test Set A. Among them, the P, R, and $mAP_{0.5}$ of the model on Test Set B were 3.3%, 16.3%, and 10.5% lower than Test Set A, respectively. The data indicates that the model's missed segmentation rate increased, and it also reflected the difficulty in small-scale object segmentation. In addition, the model's segmentation performance for stalks was lower than that for lotus pods. The $mAP_{0.5}$ for stalks was 10.9% lower than that of the lotus pods. This is because the stalk object was much smaller than the lotus pods, so it occupied fewer pixels in the image, making it more difficult to be detected and segmented.

Figure 12a,b show the representative segmentation effects of LPSS-YOLOv5 on Test Sets A and B, respectively. Both medium-large scale lotus pods and stalks were effectively segmented. The generated masks fit the objects' boundaries well. In the presence of multiple lotus pods, the model could accurately segment each object (Figure 12(a3)). In addition, the model could still achieve accurate segmentation even if the lotus pod was partially obscured by surrounding lotus leaves, as shown in Figure 12(a4).

Figure 12b shows the segmentation effects of the model on Test Set B. The model successfully segmented small-scale lotus pods and stalks objects in the actual growth environment, even though their size accounts for a very small proportion in the image. In addition, it is worth mentioning that similar objects such as leave stalks in the image were not misidentified.

The above results indicate that LPSS-YOLOv5 could effectively and robustly segment lotus pods and stalks with various scales, which could meet the robotic harvesting task requirements of lotus pods in actual planting environments.

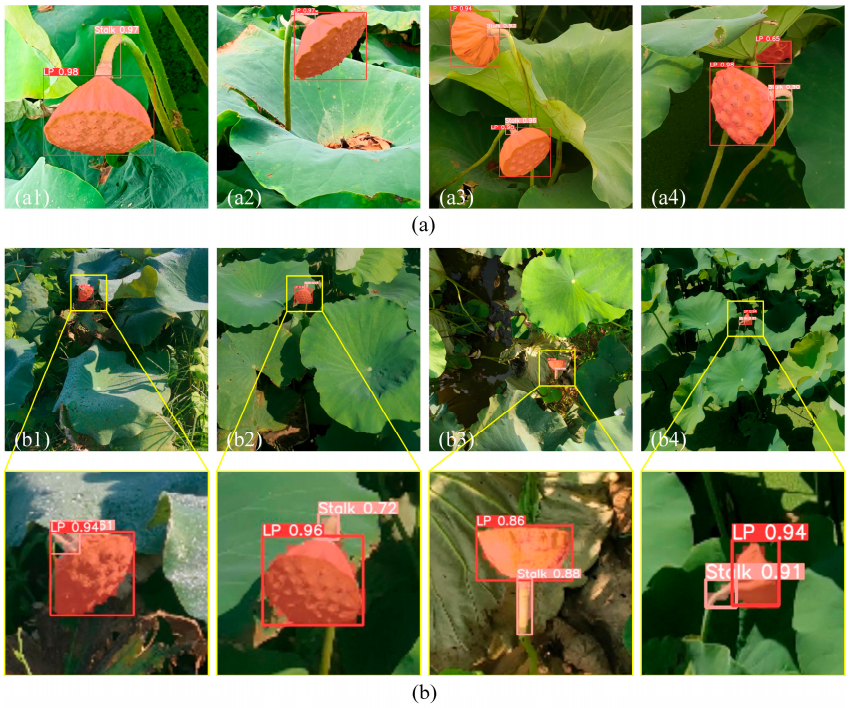

Figure 12. Segmentation effect of LPSS-YOLOv5 model on the test sets. (**a**) Test Set A. (**b**) Test Set B.

3.2. Ablation Experiment

An ablation experiment was conducted to verify the contribution of the improvement measures mentioned in Section 2.3 on the performance of the LPSS-YOLOv5 model. First, two comparative models, YOLOv5-CBAM and YOLOv5-AFL (Adjustment of the Feature Layers), were established based on YOLOv5 v7.0. YOLOv5-CBAM was built by introducing the CBAM attention mechanism into the backbone and neck networks. YOLOv5-AFL was established by adding a 160 × 160 small object detection layer and removing the original 20 × 20 large object detection layer.

Table 4 lists the results of the ablation experiment of each model on Test Sets A and B. The $mAP_{0.5}$ achieved with YOLOv5 v7.0, YOLOv5-CBAM, YOLOv5-AFL, and LPSS-YOLOv5 on Test Set A were 99.4%, 99.4%, 99.0%, and 99.3%, respectively. This means that each improvement measure had a lesser impact on the model's segmentation accuracy for medium-large objects, and each model achieved a very high segmentation accuracy. The $mAP_{0.5}$ achieved with YOLOv5 v7.0, YOLOv5-CBAM, YOLOv5-AFL, and LPSS-YOLOv5 on Test Set B were 86.2%, 87.0%, 88.3%, and 88.8%, respectively. The results of the three improved models were higher than those of the original YOLOv5 v7.0. The results indicate that both the abovementioned improvement measures play a role in improving the segmentation performance of the model for small-scale objects. Among them, the effect of introducing the two improvement measures simultaneously was more effective, obtaining a 2.6% increase in $mAP_{0.5}$.

Table 4. Test results of the ablation experiment.

Model	Add CBAM	Change Layers	Evaluation Criteria (Mask)						Parameters
			P (%)		R (%)		$mAP_{0.5}$ (%)		
			Test Set A	Test Set B	Test Set A	Test Set B	Test Set A	Test Set B	
YOLOv5 v7.0	✗	✗	97.6	91.1	99.1	82.6	99.4	86.2	7,401,119
YOLOv5-CBAM	✓	✗	98.2	92.4	98.6	81.8	99.4	87.0	7,467,631
YOLOv5-AFL	✗	✓	96.8	93.3	97.9	83.6	99.0	88.3	5,645,727
LPSS-YOLOv5	✓	✓	96.7	93.4	99.3	83.0	99.3	88.8	5,705,041

On the other hand, the model improvement affected the parameter amount of the model. The introduction of the CBAM module resulted in an increase in the number of model parameters from 7,401,119 to 7,467,631, indicating that it increased the model's complexity but to a limited extent. Adding a 160 × 160 detection layer to the network while removing the original 20 × 20 detection layer significantly reduced the model parameters from 7,401,119 to 5,645,727. This is because compared with the feature map of the 20 × 20 detection layer, the 160 × 160 feature map has fewer deep features, which reduces the model's complexity. Therefore, after adopting the above two improvement measures, the final LPSS-YOLOv5 model achieved a significant reduction in parameters compared to the original YOLOv5 v7.0 model, from 7,401,119 to 5,705,041, which means that better deployment performance could be obtained.

Table 5 shows the test results of each category in the ablation experiment. On Test Set B, the $AP_{0.5}$ values of YOLOv5-CBAM for lotus pod and stalk categories were 94.4% and 79.6%, which were higher than the 94.0% and 78.3% achieved through the original YOLOv5 v7.0 model, respectively. This verifies that introducing the CBAM enhances the model's feature extraction ability for lotus pods and stalks in complex environments. The YOLOv5-AFL model achieved $AP_{0.5}$ values of 93.8% and 82.7% for the lotus pod and stalk categories, respectively. Compared to the original YOLOv5 v7.0 model, it showed a little decrease in lotus pod segmentation. However, it achieved a 4.4% increase in the segmentation of lotus stalks. This verifies that adjusting the multi-scale feature layer structure effectively improved the model's segmentation performance for small-scale lotus stalks. With the introduction of both the improvement measures, the LPSS-YOLOv5 model achieved an $AP_{0.5}$ value of 83.3% for lotus stalk segmentation, which was a 5% increase compared to the YOLOv5 v7.0 model.

Table 5. Test results of each category in the ablation experiment.

Model	Dataset	Label	Evaluation Criteria (Mask)		
			P (%)	R (%)	$AP_{0.5}$ (%)
YOLOv5 v7.0	Test Set A	LP	97.7	100	99.5
		Stalk	97.5	98.2	99.2
	Test Set B	LP	96.8	90.1	94.0
		Stalk	85.5	75.1	78.3
YOLOv5-CBAM	Test Set A	LP	97.9	99.8	99.5
		Stalk	98.5	97.4	99.3
	Test Set B	LP	97.1	88.7	94.4
		Stalk	87.7	75.0	79.6
YOLOv5-AFL	Test Set A	LP	95.8	99.8	99.0
		Stalk	97.9	96.1	98.9
	Test Set B	LP	95.5	89.2	93.8
		Stalk	91.1	78.0	82.7
LPSS-YOLOv5	Test Set A	LP	95.6	99.8	99.3
		Stalk	97.9	98.8	99.3
	Test Set B	LP	95.4	88.7	94.2
		Stalk	91.3	77.4	83.3

Furthermore, it can be seen from the picture of the above results (Figure 13) that the LPSS-YOLOv5 model has obtained the best comprehensive segmentation performance for small objects after being improved by the two measures at the same time.

Figure 14 shows the representative segmentation effects in the test. There are both false segmentations (marked by red circles) and missed segmentations (marked by yellow circles) in the results of the YOLOv5 v7.0 model. After introducing the CBAM, the missed segmented lotus pod could be successfully detected (Figure 14d), and the falsely detected lotus flower and lotus leaves (Figure 14b,e) were not detected with YOLOv5-CBAM. The results indicate that introducing the CBAM not only improved the model's feature extraction ability for lotus pod and stalk but also suppressed the interference of surrounding

irrelevant objects and reduced the false segmentation phenomenon. However, YOLOv5-CBAM still failed to segment a small-scale stalk and a lotus pod that was heavily occluded (Figure 14a,c).

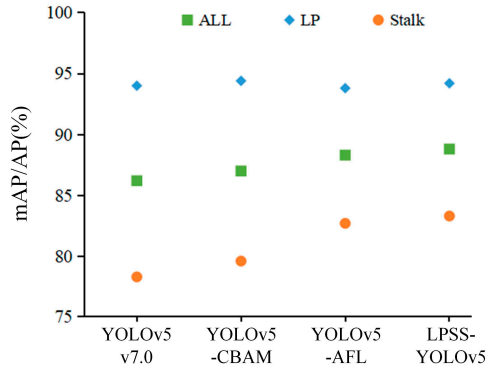

Figure 13. Test results of each category in the ablation experiment on Test Set B.

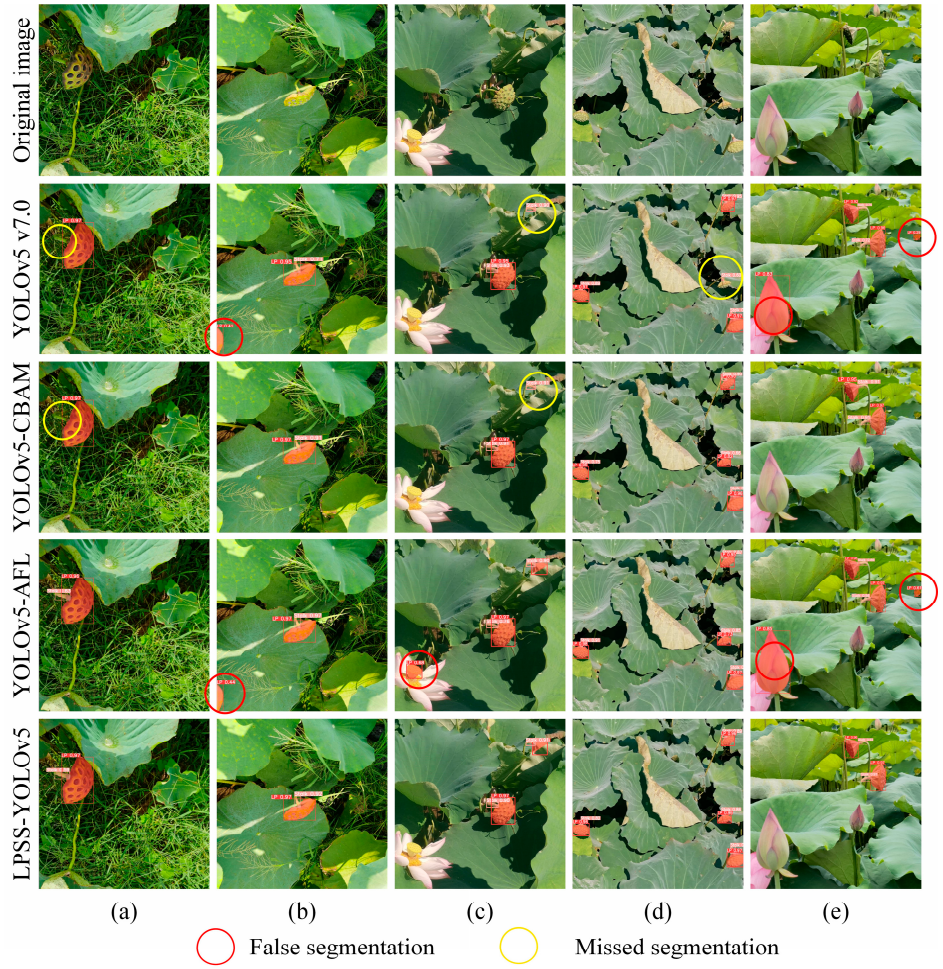

Figure 14. Segmentation effects of the ablation experiment.

On the other hand, YOLOv5-AFL achieved correct segmentation for those small stalk and occluded lotus pod objects that were not identified via the YOLOv5 v7.0 model (Figure 14a,c). However, it falsely segmented the immature lotus pod, lotus flowers, and also the leaves, which indicates that if only the AFL improvement was introduced, the model's ability to suppress the information interference of surrounding irrelevant objects was limited. In contrast, the LPSS-YOLOv5 model, which introduces both improvements,

achieved complementary advantages. The best comprehensive segmentation effect was achieved, and the abovementioned false and missed segmentation problems were avoided. In summary, according to the results of the ablation experiments, the improvements carried out on YOLOv5 v7.0 in this study have played their due roles as expected.

3.3. Performance Comparison of the Mainstream Instance Segmentation Models

To further verify the segmentation performance of the proposed model, other mainstream instance segmentation models, i.e., Mask R-CNN [34,35] and YOLACT [36], were selected and used in contrast experiments with LPSS-YOLOv5. Specifically, the same data sets were used for the comparison, and the results are shown in Table 6.

Table 6. Segmentation results of lotus pods and stalks using different mainstream instance segmentation models.

Models	Test Set	Evaluation Criteria (Mask)					FPS	Size (MB)
		mIOU (%)	P (%)	R (%)	F1-Score (%)	$mAP_{0.5}$		
YOLOv5 v7.0	Test Set A	91.8	97.6	99.1	98.3	99.4	104.1	14.7
	Test Set B	82.1	91.1	82.6	86.6	86.2		
Mask R-CNN	Test Set A	89.8	93.9	95.5	94.7	90.9	5.8	343.1
	Test Set B	82.8	68.6	33.3	44.8	33.5		
YOLACT	Test Set A	88.3	93.4	98.5	95.9	97.7	51.3	194.4
	Test Set B	83.2	78.5	78.1	78.3	74.4		
LPSS-YOLOv5	Test Set A	90.7	96.7	99.3	98.0	99.3	93.5	12.0
	Test Set B	80.0	93.4	83.0	87.9	88.8		

On Test Set A, the $mAP_{0.5}$ values of LPSS-YOLOv5 were 8.4% and 1.6% higher than the Mask R-CNN and YOLACT, respectively, and were similar to YOLOv5 v7.0. On Test Set B, the $mAP_{0.5}$ values of LPSS-YOLOv5 were 55.3%, 14.4%, and 2.6% higher than those of the Mask R-CNN, YOLACT, and YOLOv5 v7.0 models, respectively.

Figure 15 shows the representative segmentation effects of each model on the two test sets. The YOLOv5 v7.0, Mask R-CNN, and YOLACT models all have missed and false segmentation of lotus pods and stalks. In contrast, LPSS-YOLOv5 could more fully focus on the object features and pay more attention to small objects when segmenting lotus pods and stalks. Even for those objects in complex environments, the model achieved good segmentation results.

Figure 16 compares some sample predicted masks with corresponding ground truth for the four models. As shown in Figure 16a, for medium-large lotus pod and stalk objects, the difference between the predicted mask and the ground-truth mask of the above four models was small, and each model achieved effective segmentation for all objects.

As shown in Figure 16b, for the segmentation of small objects, Mask R-CNN and YOLACT achieved more smooth mask contour edges. However, they failed to segment all the objects. Combined with the performance results listed in Table 6, the problem of the high missed detection rate for small objects limits their practical application. In contrast, although the mIOU indicators reflecting mask coverage quality and contour smoothness of LPSS-YOLOv5 and YOLOv5 v7.0 were lower than that of the Mask R-CNN and YOLACT, they still achieved effective segmentation of the majority body part of the objects. Combined with the results in Table 6, on the basis of satisfying a higher recall rate, slightly lower mask coverage quality has a limited impact on the actual application effect of the model.

In terms of detection speed, LPSS-YOLOv5 achieved a detection speed of 93.5 FPS, which was much higher than Mask R-CNN's 5.8 FPS and YOLACT's 51.3 FPS but slower by 10.6 FPS compared to YOLOv5 v7.0. Although the model's improvement reduced the detection speed, it was still significantly higher than the mainstream one-stage algorithm YOLACT. In addition, the model size of LPSS-YOLOv5 was only 12 MB, which was much smaller than the 343.1 MB of Mask R-CNN and 194.4 MB of YOLACT. Therefore, LPSS-YOLOv5 is more feasible and practical to deploy on intelligent lotus pod harvesting robots than other models.

Figure 15. Segmentation effects of different instance segmentation models on the sample test images.

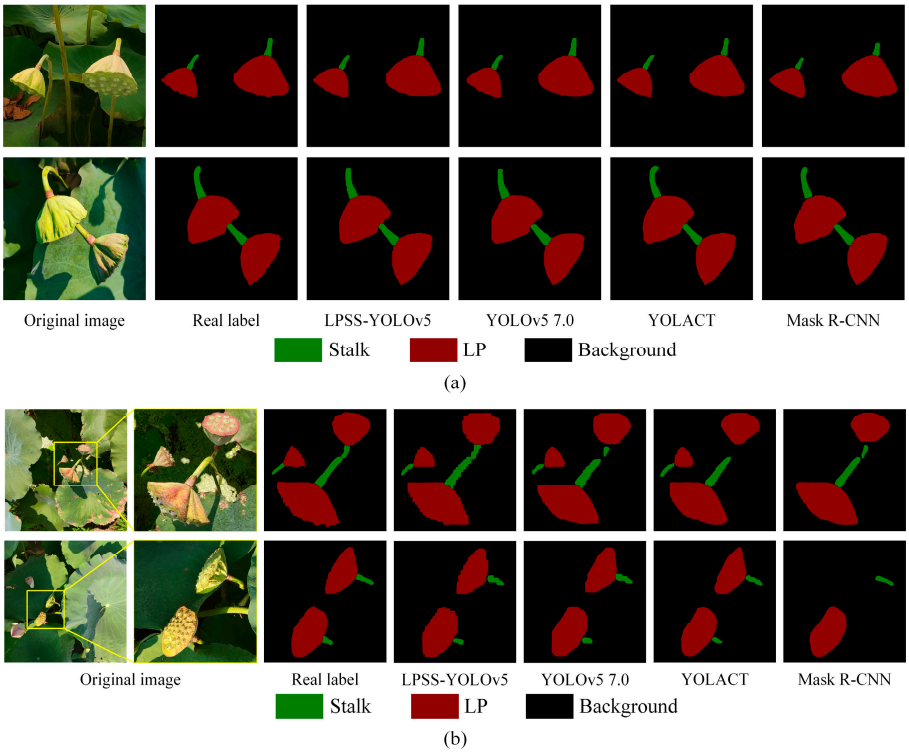

Figure 16. Segmentation effects of lotus pods and stalks obtained through different instance segmentation models. (**a**) Test Set A. (**b**) Test Set B.

Combining the results of the radar chart shown in Figure 17 and the above analysis, it is seen that the comprehensive performance of LPSS-YOLOv5 was better than other mainstream Mask R-CNN and YOLACT models in terms of segmentation quality, real-time performance, and deployability.

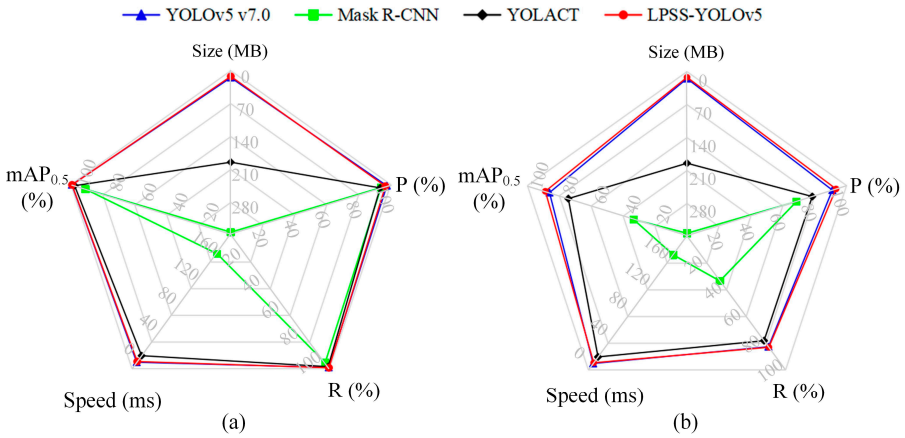

Figure 17. Performance comparison radar chart of different instance segmentation models. (**a**) Test Set A. (**b**) Test Set B.

3.4. Localization Effects of the Picking Point Based on LPSS-YOLOv5

To explore the feasibility of further using the segmentation results of LPSS-YOLOv5 for picking point localization and pod–stalk affiliation confirmation, a method for localizing stalk picking point and lotus pod's key point in the 2D image was built, and corresponding 2D localization tests were performed in this section. The implementation steps include the following: (1) Obtaining the detection box and segmentation mask of lotus pods and lotus stalks in the image using LPSS-YOLOv5. (2) Extracting the pixel data within the detection box and the mask area. (3) Calculating the centroid of each mask using Equations (7) and (8) and determining the coordinates of the pixel where the centroid is located in the image. The centroid of the stalk mask was used as the picking point, and the centroid of the lotus pod mask was used as the key point representing the individual lotus pod. The connection line between the picking point and the nearest key point was used as the pod–stalk relationship vector, which could be adopted to assist the judgment of the pose and direction of the picking end-effector.

$$x_c = \frac{\sum_{i=1}^{n} x_i f(x_i, y_i)}{\sum_{i=1}^{n} f(x_i, y_i)} \tag{7}$$

$$y_c = \frac{\sum_{i=1}^{n} y_i f(x_i, y_i)}{\sum_{i=1}^{n} f(x_i, y_i)} \tag{8}$$

where x_c, y_c represent the coordinates of the centroid pixel. n represents the total number of pixels in the detection box. x_i and y_i represent the coordinates of the ith pixel. f (x_i, y_i) is the pixel value, while the f (x_i, y_i) of the pixels in the mask is 1; otherwise it is 0.

Images with different scales and lighting conditions in the test sets were selected for testing, and the results are shown in Figure 18, respectively. It can be seen from the figures that the corresponding picking points (red) and the key points (blue) of the lotus pods under different conditions have been successfully located. In addition, when multiple lotus pods were detected (Figure 18(a2,b2,b3)), the affiliation relationship between the lotus pod and the corresponding lotus stalk could be correctly established. The test results indicate that the LPSS-YOLOv5 model is robust and could effectively support the picking point localization and the pod–stalk affiliation confirmation calculation tasks under various scales and lighting conditions.

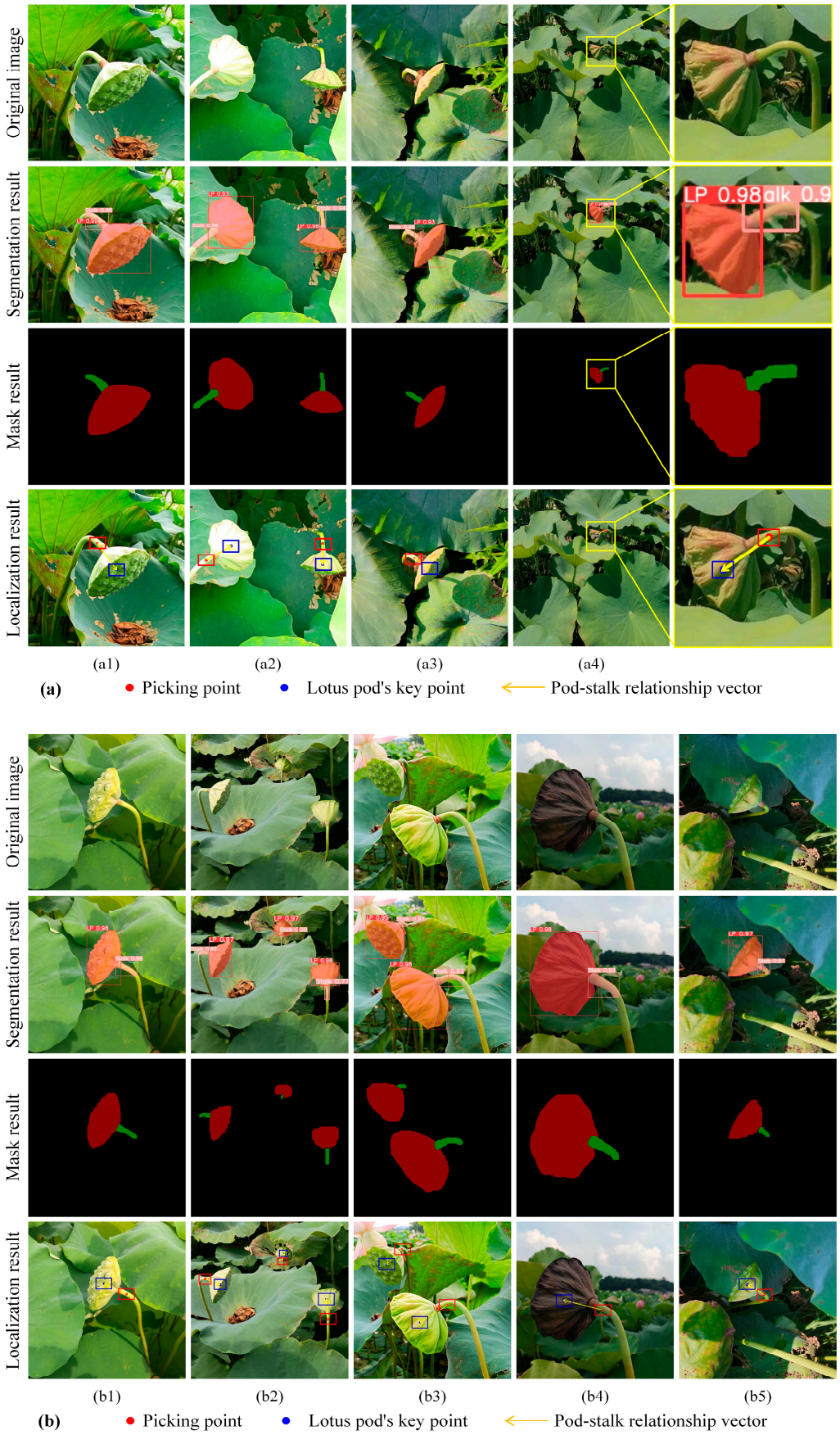

Figure 18. Two-dimensional localization effects of lotus pods and stalks under different (**a**) scales and (**b**) lighting conditions.

3.5. A 3D Localization Test Based on the Lotus Pod Harvesting Robot

Furthermore, a 3D localization test of lotus pods and stalks in the laboratory environment was conducted based on the self-developed lotus pod harvesting robot (Figure 19a). Among them, the acquisition of 3D information was performed using a side-view depth camera (Intel Realsense D435) installed on the arm of the harvesting robot (Figure 19b). The camera uses stereoscopic depth technology with a global shutter and the ideal depth range is 0.3 m~3 m. In the test, the RGB and the depth images of the sample lotus pods were captured with the camera, and the schematic diagram is shown in Figure 19c. The pixel resolution of RGB and depth images set was 1280 × 720. After the acquisition, the depth image was aligned to the color image. Then, the method described in Section 3.4 was used to obtain the 2D localization information of the picking point and the key point of the lotus pod. Subsequently, the final 3D space coordinates of the points were obtained according to the camera's intrinsic parameters and corresponding depth data.

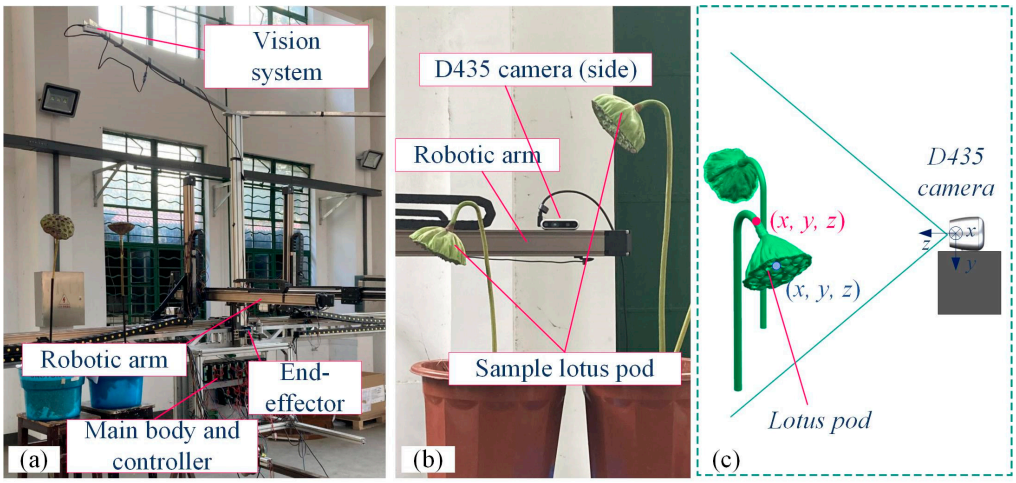

Figure 19. Three-dimensional localization test in the lab environment. (**a**) The appearance of the self-developed lotus pod harvesting robot. (**b**) Experimental scene. (**c**) Schematic diagram of the test.

The localization results are shown in Figure 20a–d. For single or multiple lotus pods, the combination of the LPSS-YOLOv5 model and the depth camera achieved effective 3D localization of the picking point and the lotus pod's key point.

In summary, the proposed LPSS-YOLOv5 model in this study showed good performance in lotus pod and stalk detection and segmentation tasks. It maintained a high segmentation speed and achieved high accuracy and robustness while reducing the complexity of the model. It could provide the location and contour information of lotus pods and stalks for lotus pod harvesting robots to perform picking operations.

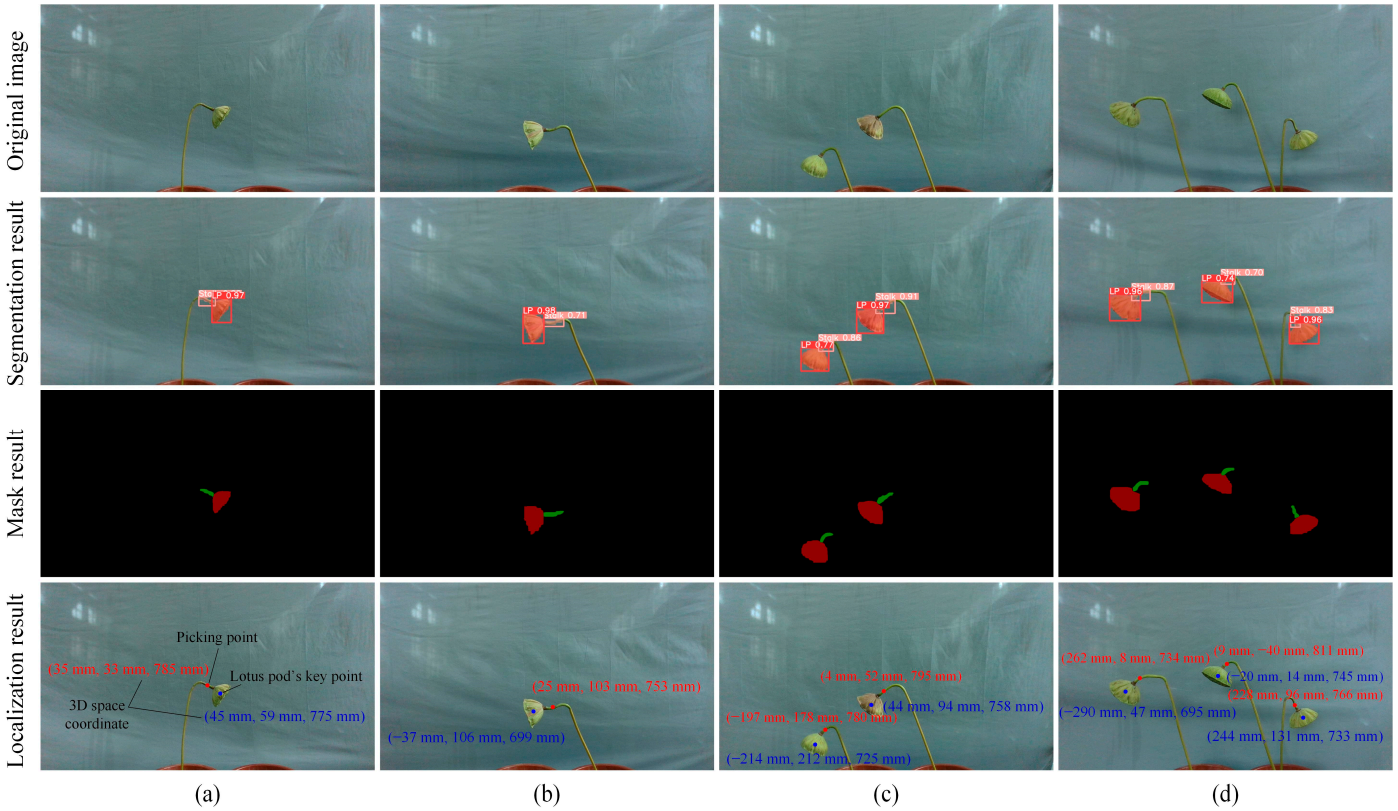

Figure 20. Three-dimensional localization effects of lotus pods and stalks in the lab environment.

4. Conclusions

To achieve accurate segmentation of lotus pod and stalk objects in the unstructured planting environment, this study proposed an instance segmentation model for lotus pods and stalks, named LPSS-YOLOv5. In the model network, the CBAM attention mechanism was introduced and the multi-scale feature layer structure of the model network was adjusted. Among them, by introducing the CBAM module, the feature extraction ability of the model for the lotus pods and stalks was improved. In addition, by adding a 160 × 160 small-scale detection layer and by removing the existing 20 × 20 large-scale detection layer, the segmentation performance of the model for small-scale lotus stalks was effectively improved. Meanwhile, the model size was reduced. The model's $mAP_{0.5}$ on the medium-large scale test set and the small-scale test set were 99.3% and 88.8%, respectively.

Compared with other mainstream instance segmentation algorithms, i.e., Mask R-CNN, YOLACT, and YOLOv5 v7.0, on the medium-large scale test set, the $mAP_{0.5}$ values of LPSS-YOLOv5 were 8.4% and 1.6% higher than the Mask R-CNN and YOLACT algorithms, respectively, and similar to YOLOv5 v7.0. On the small-scale test set, the $mAP_{0.5}$ values of LPSS-YOLOv5 were 55.3%, 14.4%, and 2.6% higher than those of Mask R-CNN, YOLACT, and YOLOv5 v7.0 algorithms, respectively. It is worth noting that the $AP_{0.5}$ for stalks achieved with LPSS-YOLOv5 was 83.3%, 5.0% higher than the YOLOv5 v7.0 model.

The LPSS-YOLOv5 achieved a detection speed of 93.5 FPS, which was much higher than Mask R-CNN's 5.8 FPS and YOLACT's 51.3 FPS. The model size of LPSS-YOLOv5 was only 12 MB, which was much smaller than the 343.1 MB of Mask R-CNN and 194.4 MB of YOLACT. The comprehensive performance of LPSS-YOLOv5 was better than other mainstream models in terms of segmentation accuracy, speed, and deployability.

Finally, a method for localizing the stalk picking point and lotus pod's key point in the 2D image was built and tested. A 3D localization test was conducted based on the self-developed lotus pod harvesting robot. The research results verified that LPSS-YOLOv5 could effectively support the picking point localization and the pod–stalk affiliation confirmation calculation.

In future work, we will explore the deployment of the LPSS-YOLOv5 model and localization method into the control system of the lotus pods harvesting robot as a basis to support the end-effector to perform multi-DOF picking.

Author Contributions: Conceptualization, A.L.; methodology, A.L. and L.M.; software, L.M.; validation, L.M. and H.C.; formal analysis, L.M. and H.C.; investigation, L.M., H.C., A.L. and J.L.; writing—original draft preparation, A.L., L.M. and H.C.; writing—review and editing, A.L. and H.C.; supervision, Q.M. and H.C.; funding acquisition, A.L. and Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (NSFC) (Grant No. 52205285, 52175255), the Natural Science Foundation of Hunan Province (Grant No. 2023JJ40629), and the special project for the construction of the Changsha-Zhuzhou-Xiangtan National Independent Innovation Demonstration Zone (Grant No. ZD-ZD20211004).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Wang, M.; Hu, W.-J.; Wang, Q.-H.; Yang, B.-Y.; Kuang, H.-X. Extraction, purification, structural characteristics, biological activities, and application of the polysaccharides from *Nelumbo nucifera* Gaertn. (lotus): A review. *Int. J. Biol. Macromol.* **2023**, *226*, 562–579. [CrossRef] [PubMed]
2. Zhang, Y.; Xu, Y.; Wang, Q.; Zhang, J.; Dai, X.; Miao, S.; Lu, X. The antioxidant capacity and nutrient composition characteristics of lotus (*Nelumbo nucifera* Gaertn.) seed juice and their relationship with color at different storage temperatures. *Food Chem. X* **2023**, *18*, 100669. [CrossRef] [PubMed]
3. Lei, Y.; Zhang, Y.; Wang, Q.; Zheng, B.; Miao, S.; Lu, X. Structural characterization and in vitro analysis of the prebiotic activity of oligosaccharides from lotus (*Nelumbo nucifera* Gaertn.) seeds. *Food Chem.* **2022**, *388*, 133045. [CrossRef] [PubMed]
4. Punia Bangar, S.; Dunno, K.; Kumar, M.; Mostafa, H.; Maqsood, S. A comprehensive review on lotus seeds (*Nelumbo nucifera* Gaertn.): Nutritional composition, health-related bioactive properties, and industrial applications. *J. Funct. Foods* **2022**, *89*, 104937. [CrossRef]
5. Sun, H.; Liu, Y.; Ma, J.; Wang, Y.; Song, H.; Li, J.; Deng, X.; Yang, D.; Liu, J.; Zhang, M.; et al. Transcriptome analysis provides strategies for postharvest lotus seeds preservation. *Postharvest Biol. Technol.* **2021**, *179*, 111583. [CrossRef]
6. Rong, J.; Wang, P.; Wang, T.; Hu, L.; Yuan, T. Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots. *Comput. Electron. Agric.* **2022**, *202*, 107430. [CrossRef]
7. Kim, J.; Pyo, H.; Jang, I.; Kang, J.; Ju, B.; Ko, K. Tomato harvesting robotic system based on Deep-ToMaToS: Deep learning network using transformation loss for 6D pose estimation of maturity classified tomatoes with side-stem. *Comput. Electron. Agric.* **2022**, *201*, 107300. [CrossRef]
8. Fu, L.; Wu, F.; Zou, X.; Jiang, Y.; Lin, J.; Yang, Z.; Duan, J. Fast detection of banana bunches and stalks in the natural environment based on deep learning. *Comput. Electron. Agric.* **2022**, *194*, 106800. [CrossRef]
9. Hu, G.; Chen, C.; Chen, J.; Sun, L.; Sugirbay, A.; Chen, Y.; Jin, H.; Zhang, S.; Bu, L. Simplified 4-DOF manipulator for rapid robotic apple harvesting. *Comput. Electron. Agric.* **2022**, *199*, 107177. [CrossRef]
10. Wang, D.; He, D. Fusion of Mask RCNN and attention mechanism for instance segmentation of apples under complex background. *Comput. Electron. Agric.* **2022**, *196*, 106864. [CrossRef]
11. Septiarini, A.; Hamdani, H.; Hatta, H.R.; Anwar, K. Automatic image segmentation of oil palm fruits by applying the contour-based approach. *Sci. Hortic.* **2020**, *261*, 108939. [CrossRef]
12. Linker, R.; Cohen, O.; Naor, A. Determination of the number of green apples in RGB images recorded in orchards. *Comput. Electron. Agric.* **2012**, *81*, 45–57. [CrossRef]
13. Fan, P.; Lang, G.; Yan, B.; Lei, X.; Guo, P.; Liu, Z.; Yang, F. A Method of Segmenting Apples Based on Gray-Centered RGB Color Space. *Remote Sens.* **2021**, *13*, 1211. [CrossRef]
14. Lu, J.; Xiang, J.; Liu, T.; Gao, Z.; Liao, M. Sichuan Pepper Recognition in Complex Environments: A Comparison Study of Traditional Segmentation versus Deep Learning Methods. *Agriculture* **2022**, *12*, 1631. [CrossRef]
15. Liu, X.; Jia, W.; Ruan, C.; Zhao, D.; Gu, Y.; Chen, W. The recognition of apple fruits in plastic bags based on block classification. *Precis. Agric.* **2018**, *19*, 735–749. [CrossRef]
16. Zheng, C.; Chen, P.; Pang, J.; Yang, X.; Chen, C.; Tu, S.; Xue, Y. A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard. *Biosyst. Eng.* **2021**, *206*, 32–54. [CrossRef]
17. Pérez-Borrero, I.; Marín-Santos, D.; Gegúndez-Arias, M.E.; Cortés-Ancos, E. A fast and accurate deep learning method for strawberry instance segmentation. *Comput. Electron. Agric.* **2020**, *178*, 105736. [CrossRef]

18. Nasiri, A.; Omid, M.; Taheri-Garavand, A.; Jafari, A. Deep learning-based precision agriculture through weed recognition in sugar beet fields. *Sustain. Comput. Inform. Syst.* **2022**, *35*, 100759. [CrossRef]
19. Azizi, A.; Abbaspour-Gilandeh, Y.; Vannier, E.; Dusséaux, R.; Mseri-Gundoshmian, T.; Moghaddam, H.A. Semantic segmentation: A modern approach for identifying soil clods in precision farming. *Biosyst. Eng.* **2020**, *196*, 172–182. [CrossRef]
20. Gu, W.; Bai, S.; Kong, L. A review on 2D instance segmentation based on deep neural networks. *Image Vis. Comput.* **2022**, *120*, 104401. [CrossRef]
21. Hussain, M.; He, L.; Schupp, J.; Lyons, D.; Heinemann, P. Green fruit segmentation and orientation estimation for robotic green fruit thinning of apples. *Comput. Electron. Agric.* **2023**, *207*, 107734. [CrossRef]
22. Gené-Mola, J.; Sanz-Cortiella, R.; Rosell-Polo, J.R.; Morros, J.-R.; Ruiz-Hidalgo, J.; Vilaplana, V.; Gregorio, E. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Comput. Electron. Agric.* **2020**, *169*, 105165. [CrossRef]
23. Li, Y.; Wang, Y.; Xu, D.; Zhang, J.; Wen, J. An Improved Mask RCNN Model for Segmentation of 'Kyoho' (Vitis labruscana) Grape Bunch and Detection of Its Maturity Level. *Agriculture* **2023**, *13*, 914. [CrossRef]
24. Zhou, J.; Zeng, S.; Chen, Y.; Kang, Z.; Li, H.; Sheng, Z. A Method of Polished Rice Image Segmentation Based on YO-LACTS for Quality Detection. *Agriculture* **2023**, *13*, 182. [CrossRef]
25. Xu, P.; Fang, N.; Liu, N.; Lin, F.; Yang, S.; Ning, J. Visual recognition of cherry tomatoes in plant factory based on improved deep instance segmentation. *Comput. Electron. Agric.* **2022**, *197*, 106991. [CrossRef]
26. Jia, W.; Zhang, Z.; Shao, W.; Hou, S.; Ji, Z.; Liu, G.; Yin, X. FoveaMask: A fast and accurate deep learning model for green fruit instance segmentation. *Comput. Electron. Agric.* **2021**, *191*, 106488. [CrossRef]
27. Liu, M.; Jia, W.; Wang, Z.; Niu, Y.; Yang, X.; Ruan, C. An accurate detection and segmentation model of obscured green fruits. *Comput. Electron. Agric.* **2022**, *197*, 106984. [CrossRef]
28. Li, Y.; Feng, Q.; Liu, C.; Xiong, Z.; Sun, Y.; Xie, F.; Li, T.; Zhao, C. MTA-YOLACT: Multitask-aware network on fruit bunch identification for cherry tomato robotic harvesting. *Eur. J. Agron.* **2023**, *146*, 126812. [CrossRef]
29. Zhong, Z.; Xiong, J.; Zheng, Z.; Liu, B.; Liao, S.; Huo, Z.; Yang, Z. A method for litchi picking points calculation in natural environment based on main fruit bearing branch detection. *Comput. Electron. Agric.* **2021**, *189*, 106398. [CrossRef]
30. Yang, C.H.; Xiong, L.Y.; Wang, Z.; Wang, Y.; Shi, G.; Kuremot, T.; Zhao, W.H.; Yang, Y. Integrated detection of citrus fruits and branches using a convolutional neural network. *Comput. Electron. Agric.* **2020**, *174*, 105469. [CrossRef]
31. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Yifu, Z.; Wong, C.; Montes, D.J.Z. Ultralytics/yolov5: v7. 0-YOLOv5 SotA Realtime Instance Segmentation. 2022. Available online: https://ui.adsabs.harvard.edu/abs/2022zndo...7347926J/abstract (accessed on 19 July 2023).
32. Huang, W.; Huo, Y.; Yang, S.; Liu, M.; Li, H.; Zhang, M. Detection of Laodelphax striatellus (small brown planthopper) based on improved YOLOv5. *Comput. Electron. Agric.* **2023**, *206*, 107657. [CrossRef]
33. Li, R.; Wu, Y. Improved YOLO v5 Wheat Ear Detection Algorithm Based on Attention Mechanism. *Electronics* **2022**, *11*, 1673. [CrossRef]
34. Yuan, L.; Qiu, Z. Mask-RCNN with spatial attention for pedestrian segmentation in cyber–physical systems. *Comput. Commun.* **2021**, *180*, 109–114. [CrossRef]
35. Qu, X.; Wang, J.; Wang, X.; Hu, Y.; Zeng, T.; Tan, T. Gravelly soil uniformity identification based on the optimized Mask R-CNN model. *Expert Syst. Appl.* **2023**, *212*, 118837. [CrossRef]
36. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. Yolact: Real-time instance segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9157–9166.