YOLOv5-lotus: an Efficient Object Detection Method for Lotus Seed Pods in a Natural Environment

Jie Ma[a], Ange Lu[b], Chen Chen[a], Xiandong Ma[c], Qiucheng Ma[b]

a. School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, Jiangsu, China

b. School of Mechanical Engineering, Xiangtan University, Xiangtan, Hunan, China

c. Engineering Department, Lancaster University, Lancaster, United Kingdom.

Abstract

Accurate detection of lotus seed pods in a nature environment is essential for agronomic applications for automated harvesting and yield mapping. Traditional detection methods are based on grower's experience, which is inefficient for the large-scale production. To improve the efficiency of harvesting lotus seed pods, this study proposes a YOLOv5-lotus method to effectively detect overripe lotus seed pods. The lotus seed pod image dataset is firstly created. An improved YOLOv5 network model based on coordinate attention (CA) module is then presented, namely YOLOv5-lotus model, where CA module is developed to strengthen the model inter-channel relationships and capture long-range dependencies with precise positional information, thus improving the detection accuracy of the algorithm. In order to reveal the feasibility and robustness of the proposed method, a number of case studies are presented on the detection of overripe lotus seed pods in various scenarios, including different poses, illuminations and degrees of occlusion. Compared with the classical YOLOv5s network, the average precision of YOLOv5-lotus model is increased by 0.7% and average detection time is reduced by 0.7ms. Compared to other state-of-the-art networks, our

detection model is able to achieve the highest average precision value, fastest efficient detection speed and higher F1 score, with the average precision being 98.3%, the recall rate being 96.3%, the precision rate being 97.3%, F1 score being 0.968 and average detection time being 19.4ms. Through case studies and comparisons, the effectiveness and superiority of the proposed approach are demonstrated. These research results can serve as a reference for the mechanization of harvesting lotus seed pods.

Keywords. Lotus seed pods; YOLOv5-lotus model; CA module; Automated harvesting

1. Introduction

Lotus seed is the fruit and seed of lotus (Nelumbo nucifera), with the thousands of years of planting history. It has a wide native distribution, ranging from India, Vietnam, Cambodia to China. In the south-east China, lotus seed as a local characteristic fruit has developed into a key agriculture industry in Fujian, Jiangxi and Hunan provinces. The cultivated area of lotus in China is exceeded 100,000 ha and the yield is about 1.2×10^8kg, ranking first in the world (Chen et al., 2021; Luo et al., 2016). The lotus seeds contain rich contents of carbohydrates, protein, B vitamins and dietary minerals (Chen et al., 2021; Punia Bangar et al., 2022). They can be used as the nutritional food and traditional Chinese medicine with a great market demand. Lotus plants are commonly cultivated in the lotus pond

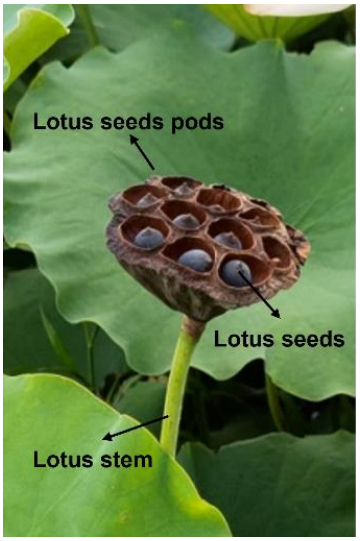


Fig. 1. Structure of lotus seed pods.

with slow-moving streams and muds. Meanwhile, plants such as lotus leaves and lotus flowers are densely distributed in the lotus pond, which brings great challenges to machine operation. Therefore, lotus seeds are mainly picked by hands and the harvesting accounts for approximately 60-70% of the labor cost. The high labor intensity, low efficiency, high labor cost and strong time restrictions of harvesting have caused great economic losses for the famers, severely limiting the development of the lotus seed industry. Thus, integrating mechanized harvesting technology into lotus seed harvesting is a promising solution. However, the first essential step to realize the automatic harvesting of lotus seeds in an unstructured environment is the detection of lotus seed pods.

(a) unripe (b) ripe (c) overripe (d) overripe

Fig. 2. Different color, texture and posture of lotus seed pods in different growth stage

In contrast to ordinary fruits that grow downward and are exposed, lotus seeds are grown in the lotus seed pod, which is a cone supported by lotus stem demonstrated in Fig. 1. Therefore, the lotus seed is harvested by harvesting the lotus seed pod and stripping the seeds. The lotus seed pods show different postures, texture and color characteristic in different growth stages of lotus seeds. Since lotus seeds in the lotus pond usually do not ripen at the same time during the harvesting season, unripe, ripe and overripe lotus seed pods viewed from same angle will differ in color, texture and

posture, as shown in Fig. 2. Regarding the overripe lotus seed pods, they present different color, texture and posture characteristics, due to the difference in sun exposure duration. They tend to bend to the side at the early stage and erect at the later stage, as shown in Fig. 2(c)-(d), respectively. Apparently, the overripe lotus seed pods have a distinct and unified appearance (i.e. the lotus seeds are black), compared to the unripe and ripe stages. After stripping the seeds from lotus seed pod, lotus seeds will be peeled, dried and the core will be removed from the seeds, before entering the market. Fig. 3 shows a series of processing procedures. The overripe lotus seeds have great economic potential in the food and medicinal market. Thus, this study focuses on the detection of lotus seed pods at overripe stage.

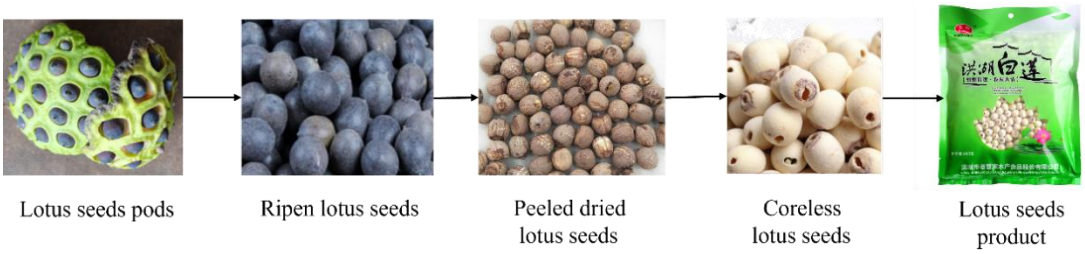


Fig. 3. The processing flow of overripe lotus seed products

Automatic computer vision systems are commonly used to detect and localize various fruits, and scholars worldwide have conducted abundant researches in this area. Generally, there are two methods, namely, digital processing and deep learning methods. The digital image processing methods utilize the color (Mim et al., 2018; Tu et al., 2018), texture (Septiarini et al., 2021) and contour (Mim et al., 2018; Yang et al., 2019) information to implement fruit recognition, which have the advantages of simplicity and low cost. The work in (Tu et al., 2018) combined red, green, blue (RGB) color space model with local-constrained linear coding to extract the color feature of passion

fruits and then detect its maturity. In (Yang et al., 2019), shape features including thickness, diameter, sphericity, aspect ratio, surface area and shape index were calculated to classify apricots. The authors in (Fashi et al., 2019) analyzed texture feature, such as skewness, elongation and entropy to classify pomegranate fruits. The authors in (Tang, 2016) utilized a series of techniques such as image preprocessing, image feature extraction, feature dimensionality reduction to establish a target recognition model for lotus seed pods. However, a major limitation of these methods is the significant error in detecting targets at occlusion, shadows, fluctuating illuminations and complex environments. The lotus seed pods grow in an unstructured environment with many obstacles, such as lotus leaf and lotus flower. Thus, digital image processing method is not the best solution.

Compared to digital image processing models, deep learning algorithms have presented a high accuracy on the object detection, which can be categorized into two classes: one-stage model and two stage model. Two stage models, such as Region Convolution Neural Network (R-CNN), Faster R-CNN, Mask R-CNN have shown a significant improvement on fruit detection (Z. Li et al., 2021; Yu et al., 2019) and segmentation (Jia et al., 2020; Wang & He, 2022; C. Zheng et al., 2021), maturity classification (Parvathi & Tamil Selvi, 2021), and yield counting (Häni et al., 2020). The authors in (Parvathi & Tamil Selvi, 2021) presented a coconut detection method based on Faster R-CNN model and a high accuracy was obtained in the environment of fluctuated illuminations, occlusion and overlap. However, the use of Faster R-CNN network possess a longer processing time and bigger model size, which is not conducive

to deployment on mobile devices. You Only Look Once (YOLO) model, as a classical one-stage network, has been widely used in fruit vision detection (C. Li et al., 2022; Shi et al., 2020). This method combines the detection, classification and localization tasks into a regression problem to simplify the network structure and reduce computation cost. The authors in (Huang et al., 2018) adopted a YOLOv2-based deep learning network to develop a lotus seed pod detector. However, this technique can only detect the front side of lotus seed pod, which is limited by the viewing angle. The authors in (Tian et al., 2019) proposed a DenseNet-fused YOLOv3 model for apple growth stage detection. This model can achieve a F1 score of 0.817 and detection time is 0.304 per frame. Along this direction, various object detection methodologies have been implemented in fruit detection using improved YOLOv3-tiny (C. Li et al., 2022), light-weighted YOLOv4 (T. Zheng et al., 2022), YOLOv4-tiny (X. Li et al., 2021; MacEachern et al., 2023). YOLOv5 is the latest detection method of the YOLO series, which is proposed by Ultralytics company (Ultralytics, 2021). It is characterized by faster speed, higher detection accuracy, and smaller model size. Many literatures have developed improved YOLOv5 models to implement fruit detection (Lv et al., 2022; Wu et al., 2022), fruit disease detection (Qi et al., 2022; Zhang et al., 2022) and precise agriculture field. The authors in (Lv et al., 2022) introduced the BiFPN-S model and ACON-C activation function in YOLOv5 model to detect apple growth forms. However, the above detection models are mainly aimed at downward-growing regular sphere. There is still no systematic study on the detection of lotus seed pods with various postures, which offer great difficulties and challenges.

This paper presents YOLOv5-based object detection model for ripen lotus seed pods by incorporating the coordinate attention (CA) module into original YOLOv5 network. CA module focuses on capturing positional information and channel-wise relationships to augment the feature representations for the network, which performs better than other attention methods, *e.g.,* squeeze-and-excitation network (SENet), convolutional block attention module (CBAM). It has been proved that the model accuracy can be improved without additional computing cost (Hou et al., 2021). By training the improved YOLOv5 model on the self-created lotus seed pod dataset, a lotus seed pod automatic detector is obtained to improve the efficiency of harvesting the lotus seed pods, enhance robustness of the detection and provide a reference for lotus seed pods detection.

The contributions of this paper are summarized as follows:

(1) The dataset of overripe lotus seed pod in the natural environment is established, including front, side and top views.

(2) A YOLOv5-lotus model based on YOLOv5 network is proposed, which is successfully trained and tested on lotus seed pod dataset. The results show that the average precision and detection time of the proposed algorithm are superior to other classical deep learning networks. Furthermore, the detection model has been applied to the self-designed experimental equipment for harvesting lotus seed pods to realize a real-time detection.

(3) The lotus seed pods are upward-growing cone fruits that present inconsistent color, shape and posture characteristic at the overripe stage, which are varied from the

common down-growing fruits. This method provides a reference for fruits with similar growth properties.

The rest of the paper is organized as follows. The introduction and related work are given in Section 1. Section 2 presents the experiment and materials, including the study location and field trails, the preparation of image datasets and experimental conditions. The proposed model consisting of YOLOv5 network and CA module is then described in Section 3. Section 4 presents the relevant experimental results and the associated discussions. Finally, Section 5 summarizes the paper.

2. Experiment and materials

2.1 Study location and field trials

The field trails were conducted at Qixing village, Huashi town, Xiangtan County, Hunan province, China (north latitude: 27°30′4.78″, east longitude: 112°42′28.59″), as presented in Fig 4. Two lotus ponds (Site A and Site B) in Qixing village having well-

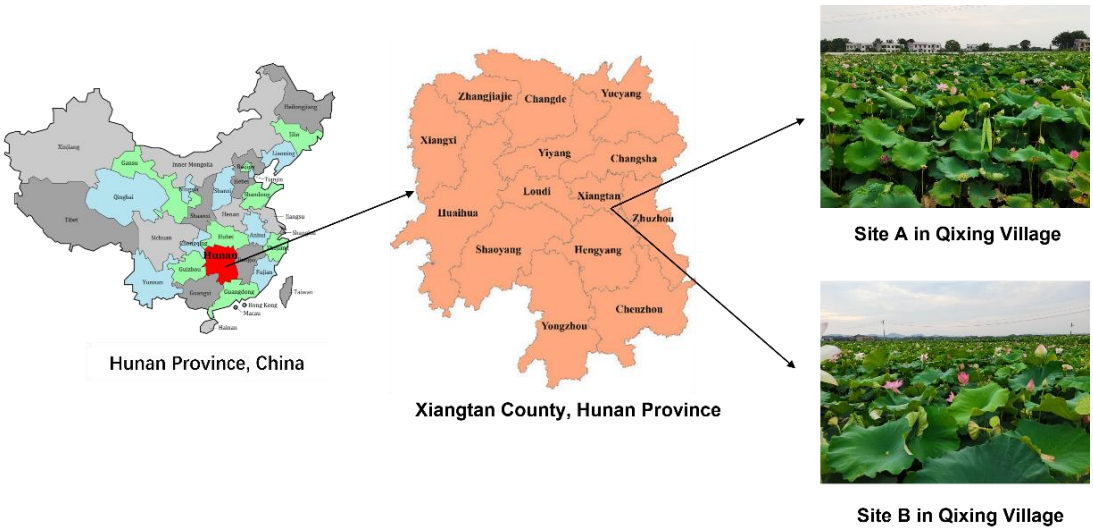


Fig. 4. The geographical location of Site A and Site B used in this study

grown lotus seed pods were selected for image acquisition. The size of Site A is 50m

by 100 m while the size of Site B is 70m by 120m. The lotus seeds planted at these two sites are Xianglian 2 and Space Lotus 36. These two lotus seeds and their lotus seed pods look similar in appearance. Images collected at these two locations ensure that there are sufficient images to evaluate robustness of the proposed model.

2.2 Image acquisition

Dataset creation of the overripe lotus seed pods is a crucial step for object detection using deep learning algorithms. In this research, image acquisition was carried out using a digital camera (Nikon-D3200), mobile phone camera (Huawei Nova 7, iPhone 11) and drone (DJI Air 2S) at Site A and Site B, from July 12 to August 15, 2021. The digital camera and mobile phone camera are used to acquire lotus seed pod images from front and side view with the distance between 30cm and 1 m. Drone is utilized to collect top-view images from 1 m to 1.2 m above the lotus seed pods, as shown in Fig. 5. As mentioned previously, the appearance and posture of overripe lotus seed pods will vary depending on the duration of sun exposure. Fig. 6(a1)-(a2), (b1)-(b2), (c1)-(c2) present front, side and top views at early and later stage of the overripe lotus seed pods, respectively. Furthermore, images were taken at different time (from morning to evening) and under various weather conditions (sunny and cloudy). Thus, the overripe lotus seed pod images with different size, color, shape, posture, brightness, degree of occlusion, and illumination were collected

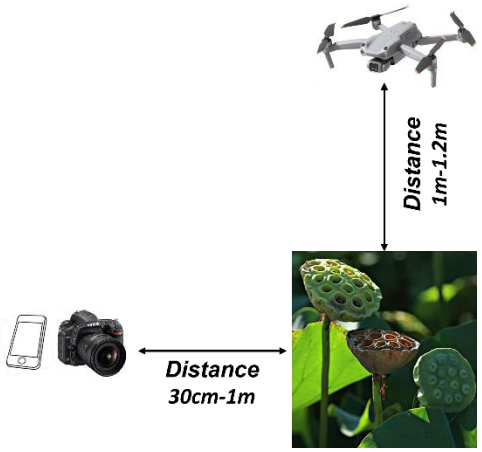


Fig. 5. Image acquisition method

to create a dataset. Overall, 1000 images of lotus seed pods at the overripe stage were used for the detection process.

(a1) (b1) (c1) (a2) (b2) (c2)

Fig. 6. Overripe lotus seeds pods. (a1)(a2) front views (b1)(b2) side views (c1)(c2) top views. (a1), (a2) and (a3) present overripe lotus seed pods at early stage, while (b1), (b2) and (b3) show the overripe lotus seed pods at later stage.

2.3 Preparation of datasets

In order to improve the quality of the experimental dataset, the collected 1000 images were then expanded to 2000 images using data augmentation techniques, such as horizontal flip, vertical flip, proportional scaling, Gaussian noise, motion blur, and random brightness. The augmented image dataset is shown in Table 1. All techniques are designed to achieve an effective generalization of the lotus seed pod detection model. After creating a dataset with 2000 images of the overripe lotus seed pods, 1700 images with 6274 fruits were randomly selected as training dataset examples that were used for establishing the deep learning models, while the remaining 300 images with 1221 fruits were assigned to the test dataset to verify the model performance.

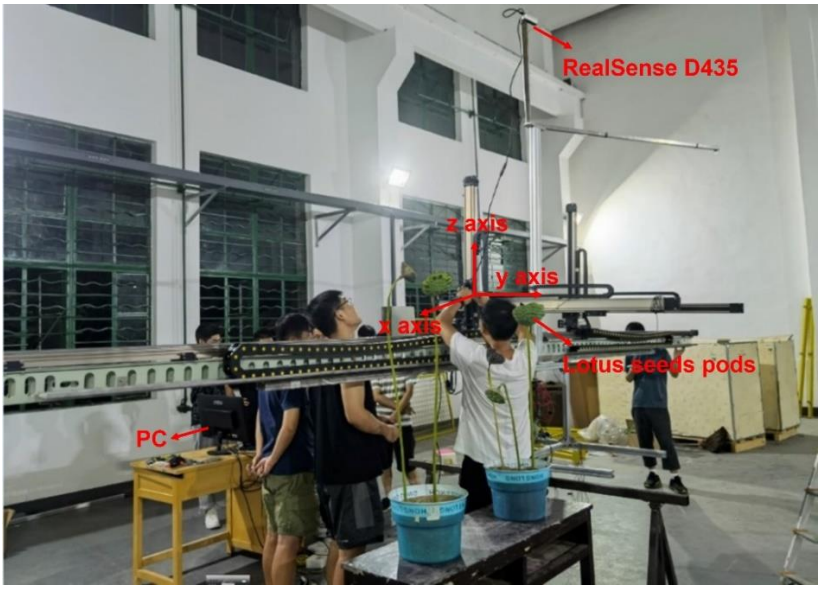
The acquired images were numbered and manually annotated using LabeImg graphical image annotation tool. Bounding boxes were drawn for the overripe lotus seed pods. The labels were stored in PASCAL VOC format or YOLO format to accommodate different algorithms.

Table 1 Number of lotus seed pod images under various data augmentation techniques.

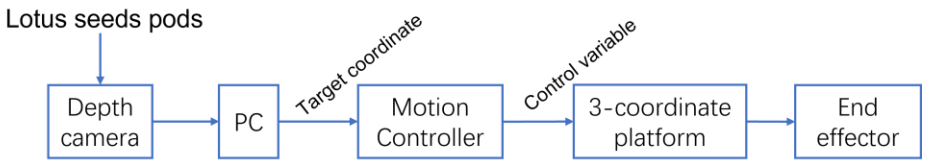
Techniques	Horizontal flip	Vertical flip	Proportional scaling	Gaussian noise	Motion blur	Random brightness	Total
Number of images	100	100	200	200	200	200	1000

2.4 Field test conditions

We have developed an experimental equipment for harvesting lotus seed pods at Xiangtan University to test the effectiveness of the detection model. The experimental setup is shown in Fig. 7(a) and the schematic diagram of experiment equipment is presented in Fig. 7(b). This harvesting equipment is made based on the three-coordinate



(a)



(b)

Fig. 7. Lotus seed pod harvesting experimental equipment: (a) experiments setup. (b) schematic diagram.

structure, which consists of a personal computer (PC), motion controller, depth camera

and three-coordinate platform. A RealSense D435 depth camera is mounted on z axis direction to acquire the lotus seed pod image from the top view. The movement of the end effector is realized by the motion controller and three-coordinate platform. PC, as a main controller, implements overripe lotus seed pod detection task and calculate their 3-dimentional coordinates. The motion controller and three-coordinate platform are used to control the movement of the end effector and perform the picking behavior.

3. Methodologies

In this section, firstly, the original YOLOv5 network is introduced. An improved YOLOv5 network incorporating the CA module is proposed to improve the performance of object detection. Then, the evaluation metrics used for comparison are explained. Finally, field test conditions and methods are given.

3.1 YOLOv5 algorithm

YOLO algorithms have the same network structure, which consists of input, backbone, neck and prediction. YOLOv5 algorithm is implemented based on YOLOv4 and YOLOv3, where tricks have been made on each component to improve the performance of the algorithm. In this paper, the YOLOv5-lotus model is proposed to improve the detection accuracy while ensuring the lightweight design of the model. YOLOv5 is selected as a benchmark model, where its structure is shown in Fig. 8. Each component of the network is described below.

The input module implements mosaic data enhancement, adaptive anchor box calculation and adaptive image scaling. Mosaic data enhancement aims to enrich the dataset and the number of small objects by randomly scaling, cropping and arranging

images followed by stitching. Adaptive anchor box module calculates the optimal anchor box size regarding different datasets. Adaptive image scaling module adaptively adds minimum black borders to the original image with different lengths and widths and uniformly scales the images to obtain a standard size.

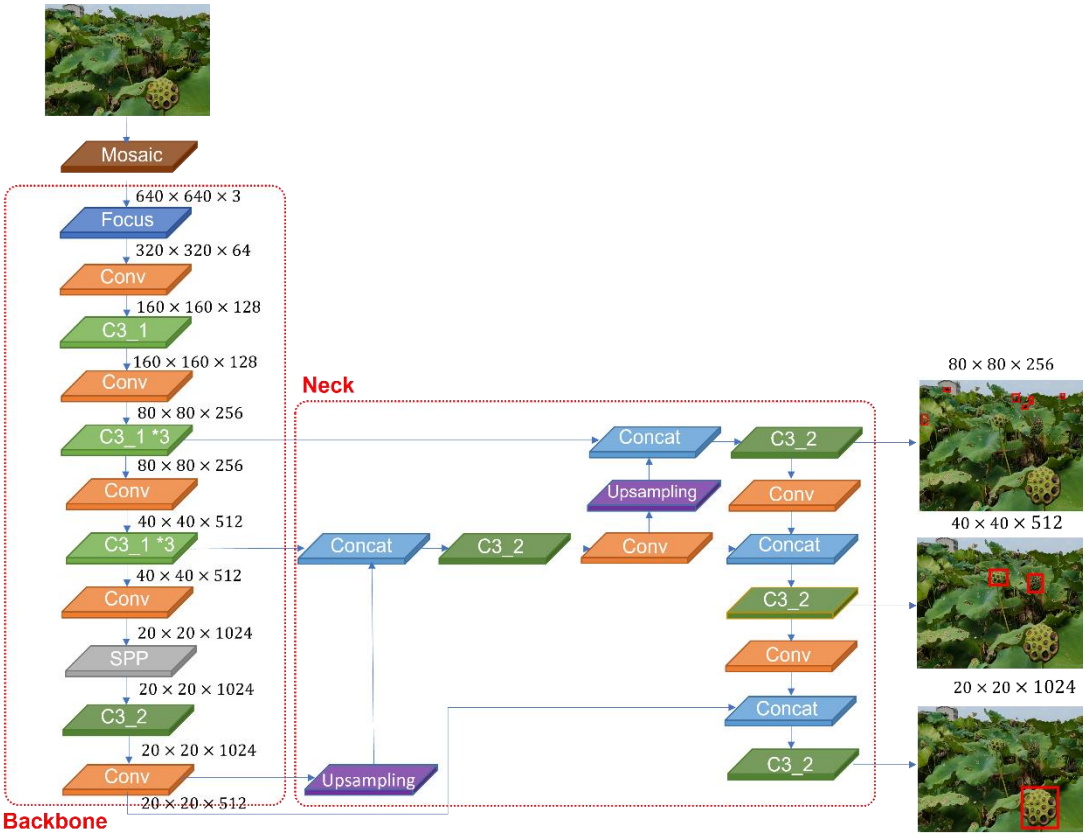


Fig. 8. YOLOv5 network schematic diagram

The backbone network of YOLOv5 is composed of Focus, Convolution block (Conv block), Cross Stage Partial (CSP) unit (C3_1 and C3_2 block) and Spatial Pyramid Pooling (SPP) block, as shown in Fig. 8. Focus module slices the input image into four pieces, which is equivalent to obtain downsampled feature map without information loss. Convolution block is a basic convolution unit of YOLOv5, which performs convolution, normalization and activation operation on the input. CSP structure is integrated into YOLOv5 network to strengthen learning ability of CNN,

remove computational bottlenecks and reduce memory costs. It divides the input feature map into two parts and merge them through a cross-stage hierarchy, in order to reduce computation cost and ensure detection accuracy. SPP module executes maximum pooling by using different kernel sizes. Then, a concatenation operation is performed to fuse features maps of different scales.

The neck network of YOLOv5 performs multi-scale feature fusion by using feature pyramid network (FPN) and pixel aggregation network (PAN). FPN structure implements downsampling operation on top feature maps and fuse lower feature maps to convey the deep sematic information. PAN structure aims to convey localization feature from the lower feature maps to the top feature maps. These two structures sufficiently fuse the multi-scale features extracted from the backbone network and jointly strengthen feature fusion ability of the neck network.

The detection network is composed of three detection layers, where the feature map with the size of 20x20, 40x40, 80x80 is used to detect small, medium and large target objects, respectively. Each detection layer generates a vector information including the probability of target object, the object score and the position of the bounding box of the target object.

3.2 Model improvements

In order to achieve lotus seed pod detection accuracy and speed, an improved YOLOv5 algorithm incorporated with the CA module is proposed. The purpose of CA module is to mitigate the loss of position information caused by 2-dimensional global pooling by embedding the coordinate information into channel attention, thereby

improving the network efficiency and reducing computational costs.

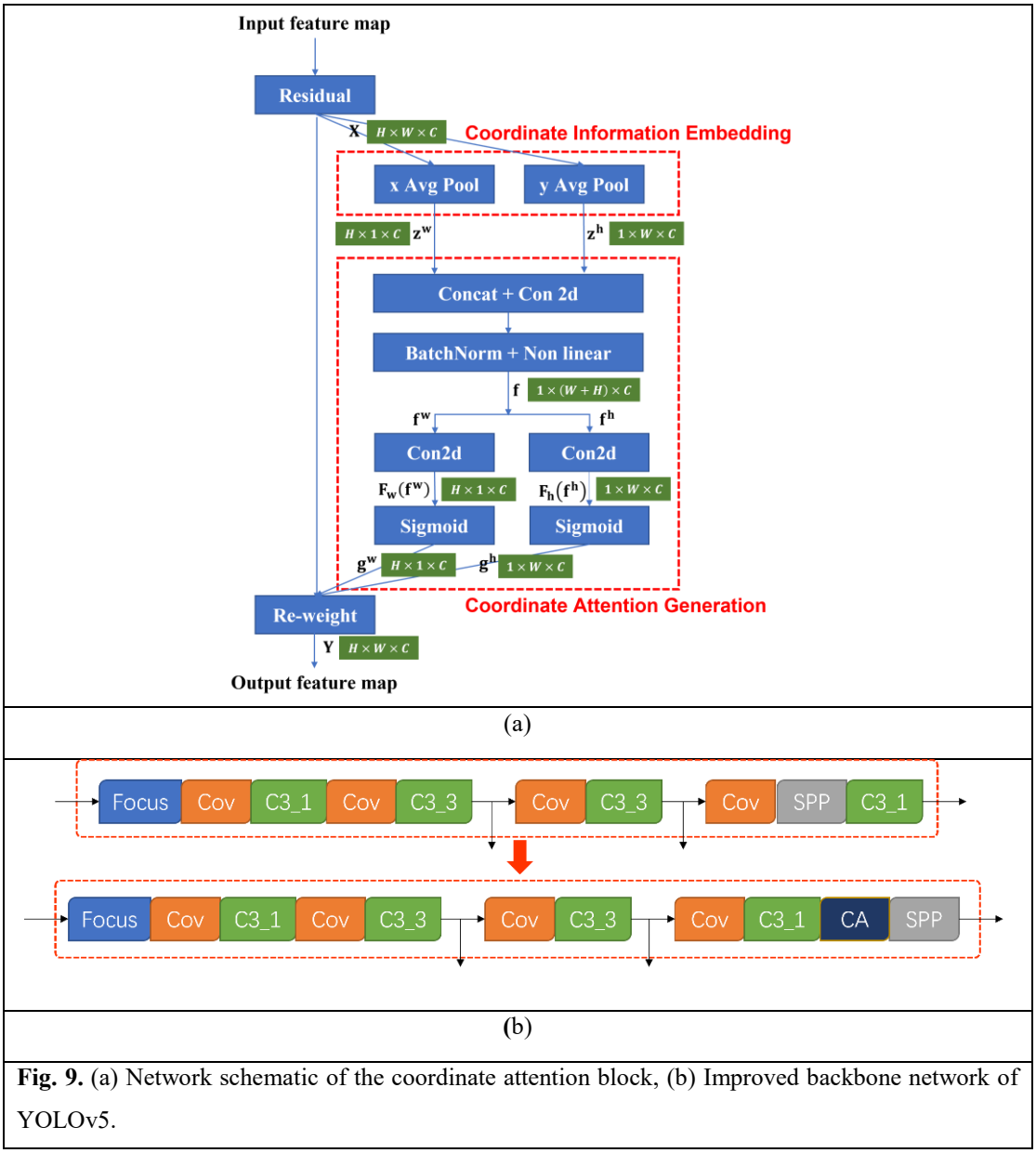


Fig. 9. (a) Network schematic of the coordinate attention block, (b) Improved backbone network of YOLOv5.

CA is a lightweight, widely used attention mechanism that not only captures cross-channel information, but also considers direction-aware and position-sensitive information. It is implemented by two steps, namely, coordinate information embedding and coordinate attention generation. Fig. 9(a) shows the network schematic of coordinate attention block, where green boxes denote the dimension of the feature tensor. Let define the input and output feature tensor as $\mathbf{X} = [\mathbf{x_1}, \dots, \mathbf{x_c}, \dots, \mathbf{x_C}]$, $\mathbf{Y} =$

$[\mathbf{y_1}, \ldots, \mathbf{y_c}, \ldots, \mathbf{y_C}]$, $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{H \times W \times C}$ where H, W and C denote the height, width and channel number of the feature map $\mathbf{X}$, respectively. Coordinate information embedding block aims to aggregate features along the horizontal coordinate and vertical coordinate by using two pooling kernels. 'X Avg Pool' and 'Y Avg Pool' refer to 1D horizontal global pooling and 1D vertical global pooling, respectively. The output of c-th channel at height h and width w can be formulated as:

$$\mathbf{z}_c{}^h(h) = \frac{1}{W} \sum_{0 \le i \le W} \mathbf{x}_c(h, i) \tag{1}$$

$$\mathbf{z}_c{}^w(w) = \frac{1}{H} \sum_{0 \le j \le H} \mathbf{x}_c(j, w) \tag{2}$$

Coordinate attention generation block aims to effectively capture inter-channel relationships and positional information. By concatenating the feature map $\mathbf{z}^h \; and \mathbf{z}^w$ and sending to a convolutional transformation function F_1, it yields:

$$\mathbf{f} = \delta\left(F_1([\mathbf{z}^h, \mathbf{z}^w])\right) \tag{3}$$

where $[.,.]$ denotes the concatenation operation along the spatial dimension, δ is a non-linear activation function and $\mathbf{f}$ is the intermediate feature map encoding the vertical direction and horizontal direction information with the size of $1 \times (W + H) \times C$. $\mathbf{f}$ can be split into two tensors $\mathbf{f}^h$and $\mathbf{f}^w$. A 2-Dimensional convolution transformation F_w and F_h are performed to separately transform $\mathbf{f}^h$ and $\mathbf{f}^w$, yielding:

$$\mathbf{g}^h = \sigma\left(F_h(\mathbf{f}^h)\right) \tag{4}$$

$$\mathbf{g}^w = \sigma\left(F_w(\mathbf{f}^w)\right) \tag{5}$$

where σ is the sigmoid activation function. $\mathbf{g}^h$ and $\mathbf{g}^w$ demonstrate the attention weights in horizontal and vertical direction. Finally, the c-th channel output of the coordinate attention $\mathbf{y}_c$ can be written as:

$$\mathbf{y}_c(i,j) = \mathbf{x}_c(i,j) \times \mathbf{g}_c{}^h(i) \times \mathbf{g}_c{}^w(j) \quad (6)$$

Through the coordinate attention block, the spatial attention in both horizontal and vertical directions is applied to the input tensor $\mathbf{x}$. This method improves the localization accuracy of the object of interest without sacrificing additional computation cost.

As demonstrated in Fig. 9(b), CA module is embedded at the end of backbone network of YOLOv5s to strengthen the coordinate information for the overripe lotus seed pod detection and improve the accuracy of image information processing.

3.3 Evaluation criteria

In order to validate the performance of improved YOLOv5s and evaluate the detection results, precision (P), recall (R), average precision (AP), F1 score ($F1$), detection time (t_r), parameter amount, floating-point operations per second (FLOPs) and model size are used as evaluation indicators. P represents the proportion of true positive samples in the positive samples predicted by the detector. R represents the proportion of positive samples predicted by the detector in terms of in total positive samples. However, P and R fail to evaluate detection accuracy directly. $AP$and $F1$ are introduced to evaluate the capability of detection network. AP represents the average precision rate in overripe lotus seed pod detection. $F1$ is the harmonic mean of precision and recall. Higher AP and $F1$ index represent higher accuracy of the detection network. t_r represents the average detection time including preprocessing time, inference time and non-max suppression time. P, R, $F1$ and AP can be calculated by Equations (7)-(10).

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{9}$$

$$AP = \int_0^1 P(R)dR \tag{10}$$

where TP denotes the number of positive samples predicted as positive, FN denotes the number of negative samples predicted as negative, FP denotes the number of negative samples predicted as positive. Intersection set IoU indicates the overlap ratio between the predicted bounding box and true bounding box. The sample is defined as TP when its IoU is greater than the set threshold 0.5. If IoU is less than the set threshold 0.5, this sample is defined as false sample FP.

4. Results and analysis of experiments

In this section, the performance of the proposed method is evaluated, compared and discussed by a number of experiments. The proposed algorithm used PyTorch deep learning framework to detect the overripe lotus seed pod images and achieved a high detection rate under various scenarios. All the training and evaluation experiments were implemented on an NVIDA GTX 3070 GPU with 8GB memory and Intel Core i7-11700K processor with 16 GB memory.

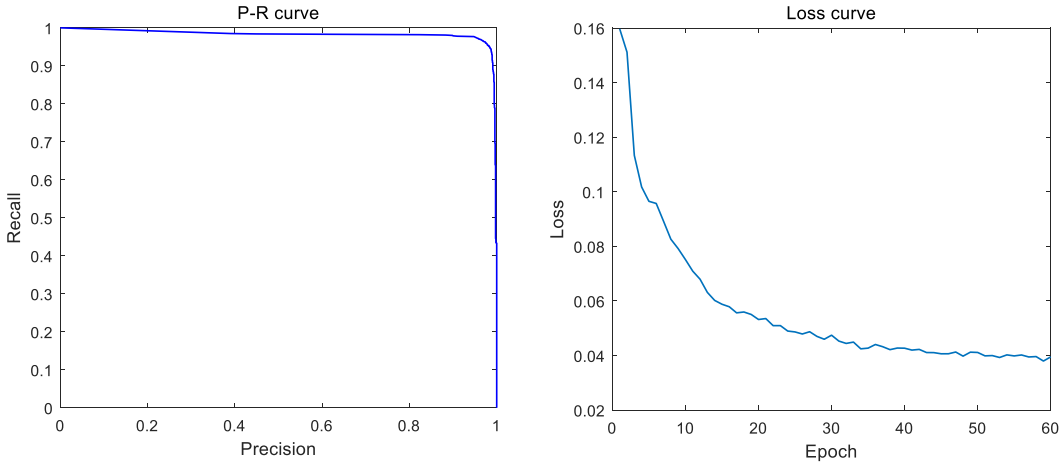
4.1 Performance of YOLOv5-lotus model

In order to verify the detection performance, the proposed YOLOv5-lotus model is trained and the recognition results of the mode on test sets were further analyzed. There are 1221 lotus seed pod targets in total in 300 test set images. The specific

evaluation results of the proposed detection model are given in Table 2, which indicates P, R, AP and F1 score of the proposed model for overripe lotus seed pods are 0.973, 0.963, 0.983, 0.968, respectively. Regarding the detection speed, YOLOv5-lotus model took an average of 19.4 ms to detect an image. The size of model weighting is 14.4MB. The precision-recall curve and loss curve are presented in Fig 10.

Table 2. Evaluation results of YOLOv5-lotus detection model on the test sets.

	Number	P	R	AP	F1	T_r	Size
YOLOv5-lotus	1221	0.973	0.963	0.983	0.968	19.4ms	14.4MB



(a) (b)

Fig. 10. P-R curve and loss curve of the YOLOv5-lotus model

As mentioned earlier, overripe lotus seed pods vary in color, shape, texture, posture, viewed from the same perspective. Thus, YOLOv5-lotus model needs to be tested against overripe lotus seed pod in different postures. Fig. 11(a)(b) and (c) show the detection results of lotus seed pods from front, back and top views, respectively. Clearly, YOLOv5-lotus model can detect all the overripe lotus seed pod targets from three views without omission.

(a) (b)

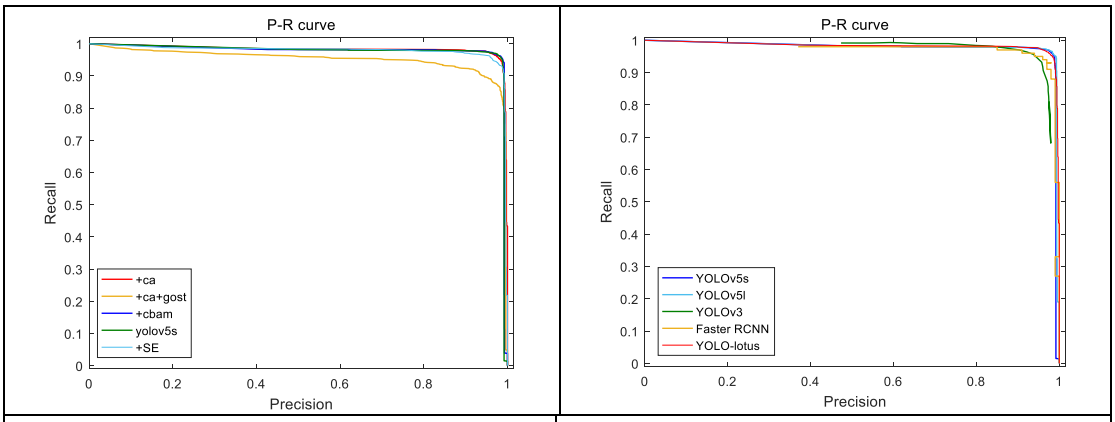
(c)

Fig. 11. Detection results of lotus seed pods in different viewing angles. (a) front side detection; (b) back side detection; (c) top view detection

4.2 Performance of the improved YOLOv5 model

Ablation experiments are performed to demonstrate that the CA module is an optimal technique to improve detection performance of overripe lotus seed pods, compared with other channel attention strategies. Five models are trained in this case, including YOLOv5s, YOLOv5s with CA module (YOLO-Lotus model), YOLOv5s with CBAM module, YOLOv5s with SE module, and YOLOv5s with CA and Ghost module. Fig. 12(a) shows the improved YOLOv5 model based on attention modules on the test set and the specific results are given in Table 3. The AP value of the proposed method (YOLO-Lotus model) is higher than those benchmark models and presents the

fastest average detection speed. It can be seen in Table 3 that by incorporating the CBAM module in YOLOv5s network, the model accuracy increases by 0.2% and the average detection processing time increased by 0.7ms, as compared with YOLOv5s. After introducing the SENet modules in YOLOv5s backbone network, the model accuracy increases by 0.3% and the detection processing time reduces by 0.9ms, as compared with YOLOv5s.



(a) (b)

Fig. 12. (a) P-R curves for improved YOLOv5 model based on attention modules; (b) P-R curves for classic detection models

Table 3. Performance measure values for YOLOv5

Model	Improve strategy	AP (0.5)	Size (MB)	Params	GFLOPs	T_r
YOLOv5s	None	97.6%	14.4	7,053,910	16.3	20.1ms
	+CA(YOLO-lotus)	98.3%	14.4	7,046,934	15.9	19.4ms
	+CBAM	97.8%	14.4	6,770,214	16.0	20.8ms
	+SE	97.9%	14.4	7,045,590	15.8	19.2ms
	+Ghost + CA	95.2%	8.1	4,220,053	9.7	20.1ms

Furthermore, incorporating the Ghost module in YOLOv5s backbone network is another effective strategy to reduce the network complexity and computation cost, by which the model size and number of model parameters are reduced by 43% and 40.2%, respectively, when compared with YOLOv5s model. However, the model accuracy decreases by 2.4%. Considering the overall results of the evaluation performance,

integrating the CA module into the YOLOv5s network is an optimal solution to developing the lotus seed pod detection model.

4.3 Comparison among the algorithms

In order to evaluate the performance of YOLOv5-Lotus model, this section compares the proposed algorithm with four classical object detection networks, i.e., YOLOv5, YOLOv3, YOLOX, and Faster R-CNN. The evaluation results of different lotus seed pod detection models are shown in Table 4 while P-R curves of each model are shown in Fig. 12(b). It can be seen that the proposed method has the highest AP, fastest detection speed and smallest model size as compared with all other methods. Compared with YOLOv5s, although the F1 score of the YOLOv5-lotus slightly decreases, the model accuracy is improved by 0.7% and detecting processing time is reduced by 0.7ms. On the other hand, YOLOv5l can also achieve a high AP value and F1 score. However, its model size is much larger than the YOLOv5-lotus.

Table 4. Test results of the benchmark models

Model	AP (0.5)	F1	P	R	T_r	Weight(M)
Our method	98.3%	0.968	0.973	0.963	19.4	14.4
YOLOv5s	97.6%	0.971	0.976	0.968	20.1	14.4
YOLOv5l	98%	0.974	0.983	0.966	22.5	93.7
YOLOv3	90.1%	0.947	0.945	0.95	21.3	234
YOLOX	90.7%	0.959	0.975	0.943	25.27	65.8
Faster RCNN	97.3%	0.92	0.866	0.973	29.8	111

Fig. 13 compares the detection performance of the proposed method and other four models in the same scenes. It is clear that YOLOv5-lotus can detect all the overripe lotus seed pods and has the highest positioning accuracy for prediction box, as shown in Fig. 13 (e). YOLOv5s and YOLOX fail to detect all targets, as presented in Fig.

13(a)(b). YOLOv3 has false detection targets, as seen in Fig. 13(c). Although Faster RCNN can detect all target fruits, its prediction boxes do not completely envelop the

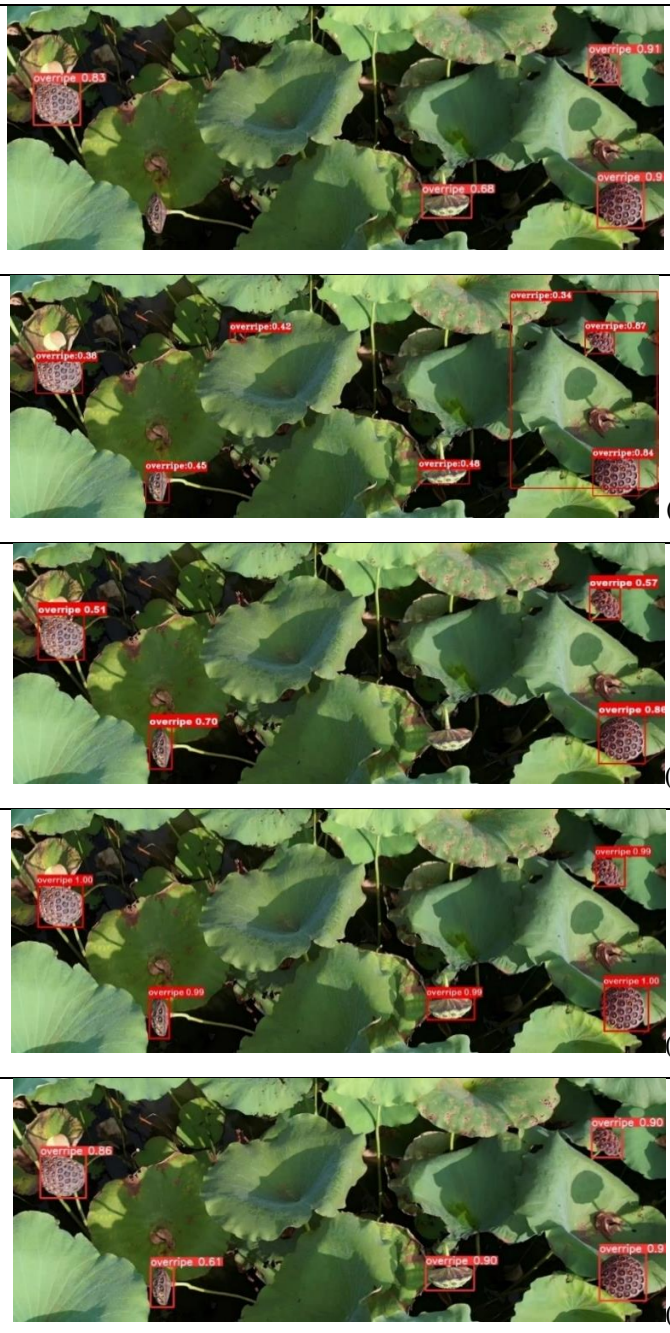


Fig. 13. Detection comparison between YOLOv5-lotus and four classic models: (a) YOLOv5s, (b) YOLOX, (c) YOLOv3, (d) Faster RCNN, (e) YOLOv5-lotus.

edge of the lotus seed pods, as shown in Fig. 13 (d).

4.4 Robustness of YOLOv5-lotus detection model

For the sake of testing the environmental adaptability of YOLOv5-lotus model,

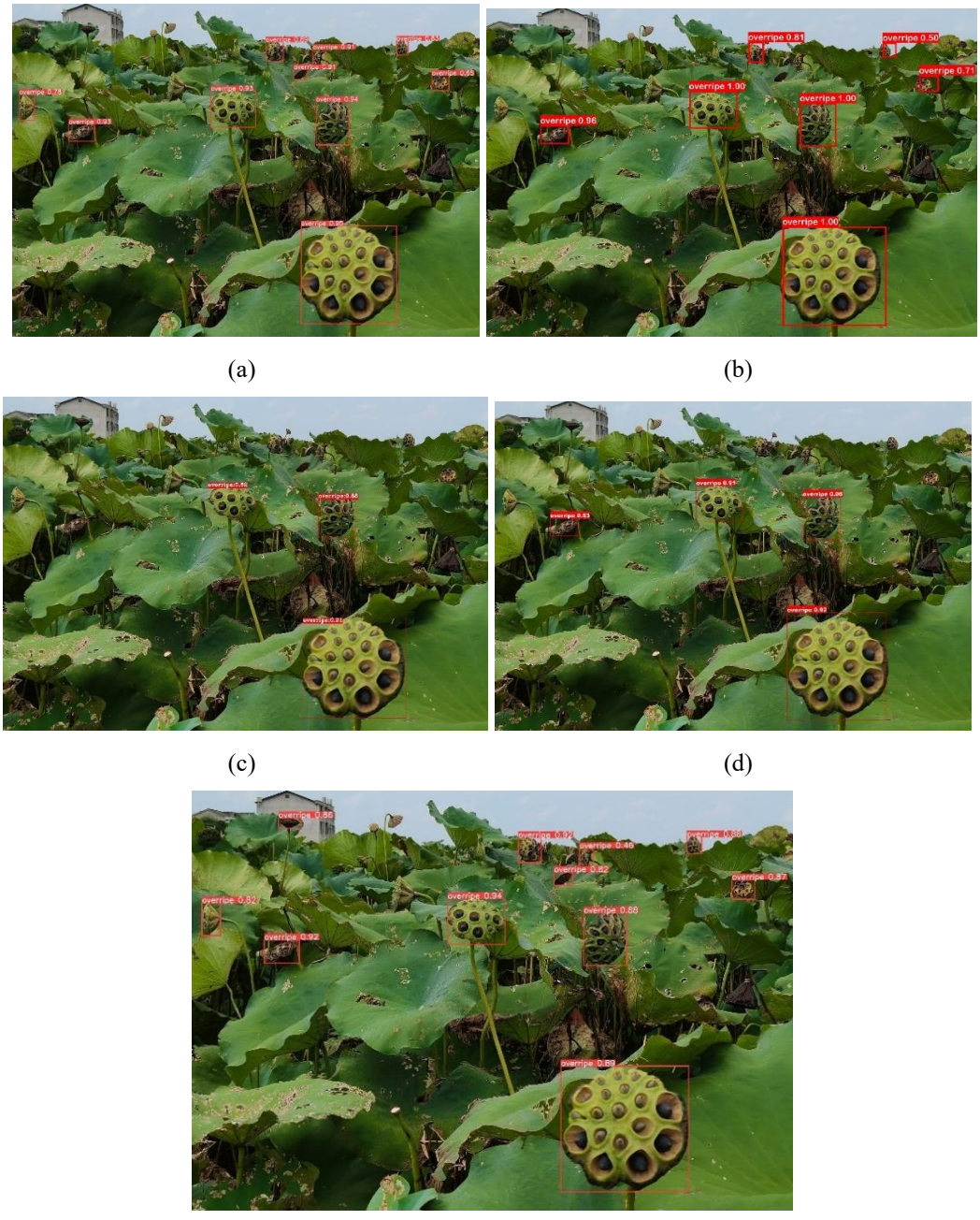
this section discusses overripe lotus seed pod detection experiments in various scenarios, such as different leave occlusion degrees, illumination intensities and distances. As shown in Fig. 14, YOLOv5-lotus model demonstrates the strong robustness in the complex nature environment of backlight, overexposure and occlusion. As shown in Fig. 14(a) and (c), when lotus seed pods are occluded by lotus leaves, YOLOv5-lotus model can detect the occluded targets with a high detection accuracy. The lotus seed pod targets occluded by the lotus leaves can be detected even in overexposure environments, as presented in the lower left corner of Fig. 14(a) and the lower right corner of Fig. 14(c).

(a) (b) (c)

Fig. 14. Detection performance of YOLOv5-lotus under different conditions: (a) occlusion, (b) backlight, (c) overexposure.

If the harvesting equipment enables to find a relatively far or small target in field of view, the harvesting efficiency can be improved. As can be seen in Fig. 15, the detection accuracy of YOLOv5-lotus for small targets and distant targets is significantly higher than other four network models. In the complex nature environment, YOLOv5-

louts model can overcome the problem of missing detection of relative distant target and achieve a high prediction frame positioning accuracy, as compared with YOLOv5s, Faster RCNN, YOLOX and YOLOv3.



(a) (b)

(c) (d)

(e)

Fig. 15. Comparison of the detection results of YOLOv5-lotus and benchmark models for small and distant targets: (a) YOLOv5s, (b) Faster RCNN, (c) YOLOX, (d) YOLOv3, and (e) YOLOv5-lotus.

Conclusions

The lotus seed pod detection is the first essential step to realize the automatic harvesting of lotus seeds in a natural environment. In this study, we collected lotus seed pod images and presented a YOLOv5-lotus model by incorporating CA technique into the YOLOv5 network. The detection results demonstrate that the improved YOLOv5-lotus model can accurately and quickly detect overripe lotus seed pods and is suitable for detection of lotus seed pods with different poses, illuminations, distances and degrees of occlusion. Experiments performed in natural environments showed that the average precision, recall rate, precision rate and F1 score of the YOLOv5-lotus model are high up to 98.3%, 96.3%, 97.3% and 0.968, respectively. The average image detection time is 19.4ms. Compared with the classical YOLOv3, YOLOv5s, Faster R-CNN and YOLOX algorithms, this model has the highest average precision, shortest detection time and best comprehensive performance.

The detection methods from this study can serve as a useful reference for further research on the detection of conical fruit growing upwards. Although the YOLOv5-lotus model proposed in this paper can efficiently detect lotus seed pods under various conditions, the current detection performance results were obtained from testing on well-defined photographs with a limited sample size. In future work, more images will be collected, and a larger dataset will be built for lotus seed pods detection by considering complex environments of the lotus pond. This will provide prerequisites for the development of automatic harvesting equipment for lotus seed pods.

Acknowledgements

This project was supported by National Key Research Foundation of China under

Grant No. 32102598 ,…,, Jiangsu University Senior Talents Start-up Fund Grant No. 5501140007 and Open Project of the Engineering Research Center of the Ministry of Education for Complex Track Processing Technology and Equipment of Xiangtan University under Grant No. FZGJ 2020-007.

References

Chen, C., Li, G., & Zhu, F. (2021). A novel starch from lotus (Nelumbo nucifera) seeds: Composition, structure, properties and modifications. *Food Hydrocolloids*, *120*, 106899. https://doi.org/https://doi.org/10.1016/j.foodhyd.2021.106899

Fashi, M., Naderloo, L., & Javadikia, H. (2019). The relationship between the appearance of pomegranate fruit and color and size of arils based on image processing. *Postharvest Biology and Technology*, *154*, 52–57. https://doi.org/https://doi.org/10.1016/j.postharvbio.2019.04.017

Häni, N., Roy, P., & Isler, V. (2020). A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics*, *37*(2), 263–282. https://doi.org/https://doi.org/10.1002/rob.21902

Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate Attention for Efficient Mobile Network Design. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350

Huang, X., Liang, Z., He, Z., & Huang, C. (2018). Research on lotus quick recognition based on YOLOv2. *Transactions of the Chinese Society of Agricultural Engineering*, *13*, 164–168.

Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., & Zheng, Y. (2020). Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Computers and Electronics in Agriculture*, *172*, 105380. https://doi.org/https://doi.org/10.1016/j.compag.2020.105380

Li, C., Lin, J., Li, B., Zhang, S., & Li, J. (2022). Partition harvesting of a column-comb litchi harvester based on 3D clustering. *Computers and Electronics in Agriculture*, *197*, 106975. https://doi.org/https://doi.org/10.1016/j.compag.2022.106975

Li, X., Pan, J., Xie, F., Zeng, J., Li, Q., Huang, X., Liu, D., & Wang, X. (2021). Fast and accurate green pepper detection in complex backgrounds via an improved Yolov4-tiny model. *Computers and Electronics in Agriculture*, *191*, 106503. https://doi.org/https://doi.org/10.1016/j.compag.2021.106503

Li, Z., Li, Y., Yang, Y., Guo, R., Yang, J., Yue, J., & Wang, Y. (2021). A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN. *Computers and Electronics in Agriculture*, *182*, 106054. https://doi.org/https://doi.org/10.1016/j.compag.2021.106054

Luo, H., Liu, X., Huang, X., Dai, X., & Zhang, M. (2016). Chemical deterioration of lotus seeds during storage: Chemical deterioration of lotus seed. *Journal of Food Quality*, *39*, 496–503.

Lv, J., Xu, H., Han, Y., Lu, W., Xu, L., Rong, H., Yang, B., Zou, L., & Ma, Z. (2022). A visual identification method for the apple growth forms in the orchard. *Computers and Electronics*

in Agriculture, *197*, 106954. https://doi.org/https://doi.org/10.1016/j.compag.2022.106954

MacEachern, C. B., Esau, T. J., Schumann, A. W., Hennessy, P. J., & Zaman, Q. U. (2023). Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks. *Smart Agricultural Technology*, *3*, 100099. https://doi.org/https://doi.org/10.1016/j.atech.2022.100099

Mim, F. S., Galib, S. Md., Hasan, Md. F., & Jerin, S. A. (2018). Automatic detection of mango ripening stages – An application of information technology to botany. *Scientia Horticulturae*, *237*, 156–163. https://doi.org/https://doi.org/10.1016/j.scienta.2018.03.057

Parvathi, S., & Tamil Selvi, S. (2021). Detection of maturity stages of coconuts in complex background using Faster R-CNN model. *Biosystems Engineering*, *202*, 119–132. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2020.12.002

Punia Bangar, S., Dunno, K., Kumar, M., Mostafa, H., & Maqsood, S. (2022). A comprehensive review on lotus seeds (Nelumbo nucifera Gaertn.): Nutritional composition, health-related bioactive properties, and industrial applications. *Journal of Functional Foods*, *89*, 104937. https://doi.org/https://doi.org/10.1016/j.jff.2022.104937

Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., Tian, X., Li, M., Bao, Z., & Li, Y. (2022). An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Computers and Electronics in Agriculture*, *194*, 106780. https://doi.org/https://doi.org/10.1016/j.compag.2022.106780

Septiarini, A., Sunyoto, A., Hamdani, H., Kasim, A. A., Utaminingrum, F., & Hatta, H. R. (2021). Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features. *Scientia Horticulturae*, *286*, 110245. https://doi.org/https://doi.org/10.1016/j.scienta.2021.110245

Shi, R., Li, T., & Yamaguchi, Y. (2020). An attribution-based pruning method for real-time mango detection with YOLO network. *Computers and Electronics in Agriculture*, *169*, 105214. https://doi.org/https://doi.org/10.1016/j.compag.2020.105214

Tang, S. (2016). *The key technology of lotus recognition based on machine vision*. Jiangsu University.

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, *157*, 417–426. https://doi.org/https://doi.org/10.1016/j.compag.2019.01.012

Tu, S., Xue, Y., Zheng, C., Qi, Y., Wan, H., & Mao, L. (2018). Detection of passion fruits and maturity classification using Red-Green-Blue Depth images. *Biosystems Engineering*, *175*, 156–167. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2018.09.004

Ultralytics. (2021). *YOLOv5*. Https://Github.Com/Ultralytics/Yolov5.

Wang, D., & He, D. (2022). Fusion of Mask RCNN and attention mechanism for instance segmentation of apples under complex background. *Computers and Electronics in Agriculture*, *196*, 106864. https://doi.org/https://doi.org/10.1016/j.compag.2022.106864

Wu, F., Duan, J., Ai, P., Chen, Z., Yang, Z., & Zou, X. (2022). Rachis detection and three-dimensional localization of cut off point for vision-based banana robot. *Computers and Electronics in Agriculture*, *198*, 107079. https://doi.org/https://doi.org/10.1016/j.compag.2022.107079

Yang, X., Zhang, R., Zhai, Z., Pang, Y., & Jin, Z. (2019). Machine learning for cultivar

classification of apricots (Prunus armeniaca L.) based on shape features. *Scientia Horticulturae*, *256*, 108524. https://doi.org/https://doi.org/10.1016/j.scienta.2019.05.051

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, *163*, 104846. https://doi.org/https://doi.org/10.1016/j.compag.2019.06.001

Zhang, D.-Y., Luo, H.-S., Wang, D.-Y., Zhou, X.-G., Li, W.-F., Gu, C.-Y., Zhang, G., & He, F.-M. (2022). Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YoloV5 method. *Computers and Electronics in Agriculture*, *198*, 107086. https://doi.org/https://doi.org/10.1016/j.compag.2022.107086

Zheng, C., Chen, P., Pang, J., Yang, X., Chen, C., Tu, S., & Xue, Y. (2021). A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard. *Biosystems Engineering*, *206*, 32–54. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2021.03.012

Zheng, T., Jiang, M., Li, Y., & Feng, M. (2022). Research on tomato detection in natural environment based on RC-YOLOv4. *Computers and Electronics in Agriculture*, *198*, 107029. https://doi.org/https://doi.org/10.1016/j.compag.2022.107029