IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 May 2023, accepted 12 June 2023, date of publication 15 June 2023, date of current version 30 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3286545

RESEARCH ARTICLE

Cucumber Flower Detection Based on YOLOv5s-SE7 Within Greenhouse Environments

XIANGYING XU[1], HONGJIANG WANG[1], MINMIN MIAO[2,3,4], WEIJIAN ZHANG[1], YONGLONG ZHANG[1], HAIBO DAI[2], ZIJIAN ZHENG[2], AND XIAOXIANG ZHANG[5]

[1]College of Information Engineering, Yangzhou University, Yangzhou 225009, China
[2]College of Horticulture and Plant Protection, Yangzhou University, Yangzhou 225009, China
[3]Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou 225009, China
[4]Key Laboratory of Plant Functional Genomics of the Ministry of Education/Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Yangzhou University, Yangzhou 225009, China
[5]Lixiahe Agricultural Research Institute of Jiangsu Province, Yangzhou, Jiangsu 225007, China

Corresponding authors: Minmin Miao (mmmiao@yzu.edu.cn) and Xiaoxiang Zhang (zhngyz@126.com)

This work was supported in part by the Research and Development Foundation of Jiangsu Province, China, under Grant BE2022425; in part by the Municipal Science and Technology Plan Project of Yangzhou under Grant YZ2021150 and Grant YZ2022179; in part by the Yangzhou University Interdisciplinary Research Foundation under Grant yzuxk202008; and in part by the Natural Science Foundation of Jiangsu under Grant BK20201219.

ABSTRACT In the process of cucumber cultivation, the quantity and appearance time of cucumber flowers are important factors influencing the final yields. However, it is labor-intensive to timely record and count the flowers. An optional method is to identify and count cucumber flowers automatically using computer vision technologies based on images taken by cameras installed in greenhouses. However, there are severe problems in images taken with a large field of view in a greenhouse environment, that is, foreground-background imbalance, which renders significant gaps between the detection accuracy and the application requirements even for state-of-the-art computer vision models like Faster-RCNN, SSD and YOLO. This problem can be improved by providing specific datasets and suitable models. Hence, in this paper, two cucumber flower datasets with a wide and medium field of view in greenhouses are constructed. Four attention mechanisms: SE, CA, CBAM and SimAM, are compared and incorporated into YOLOv5s algorithm to improve the detection performance of cucumber flowers during growing states in the greenhouse. The results indicated that our improved model with a SE attention mechanism reached the highest recognition rate than other three methods. The AP@.5 value of the YOLOv5s-SE7 model reached 0.905, which was 3.5% higher than that of the benchmark model YOLOv5s. Meanwhile, it outperformed other state-of-the-art methods such as Faster-RCNN and SSD. The classification detection results of cucumber flowers in three stages, namely bud, bloom and faded flower, reached as high as 0.847 in mAP, suggesting that the proposed model had a good effect in application.

INDEX TERMS Cucumber flower, greenhouse, object detection, YOLO.

I. INTRODUCTION

Cucumber (Cucumis sativus L.) is a major vegetable crop cultivated worldwide, which is also a model system for studies on flower development owing to its diversity of floral sex types [1], [2]. In China, it is one of the main greenhouse crops with large planting areas and productions [3]. Cucumber flowers are important reproductive organs, which not only reflect the nutritional status of the plants, but also have a great influence on the total plant productivity [4]. The quantity and appearance time of flowers are usually strictly observed in the process of cucumber cultivations for timely management, so as to achieve a high yield. With the rapid development of facility agriculture in recent years, more and more intelligent systems have been put into use while planting cucumbers in greenhouses [5], [6], [7].

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

Automatic and intelligent detection of cucumber fruits, cucumber diseases and insect pests has become an important trend in cucumber cultivations [8], [9], [10], [11], [12]. With continuous innovations of machine vision technologies, more and more crop organs and observable phenotypes such as cucumber flowers can be identified automatically, which greatly reduces the labor cost and improves the efficiency of field work [13], [14].

At present, there has been a considerable amount of research on the recognition of plant flowers through computer vision technologies, in most of which classification models are used to detect a specific flower from hundreds of plants or to detect different stages of one kind of flower. For example, Dias et al. [15] proposed a refined CNN-based network to perform automated semantic segment of fruit flowers with images collected in orchards on sunny or overcast days. Cheng and Zhang [16] presented an end-to-end flower detection method with a deep convolutional neural network, which could be integrated into a mobile device to detect more than 100 categories of flowers. Ye et al. [17] designed a polyphyletic loss function to recognize and detect occlusive litchi flowers, which had an inflorescence of corymbose cyme. Tian et al. [18] proposed an improved R-CNN model to segment apple flowers in different growth stages; Johanna et al. [19] used a light-weight convolutional neural network to recognize and classify cactus flowers in complex scenes.

Despite these achievements, there has been little research reported on cucumber flower detection in an environment of facility greenhouses. The challenges may primarily come from two aspects: datasets and models.

On the one hand, there are few public datasets on cucumber flowers. It is time-consuming and labor-intensive to build large-scale annotated datasets from scratch for all the novel categories. It is more difficult to obtain the images of greenhouse cucumbers than those of field crops because of the limited space and planting density. For example, a camera can be set up on a high support rod for imaging in a field while avoiding occlusion [20], but the support rod cannot be set up high enough in an ordinary plastic greenhouse. Furthermore, some cucumber flowers are at a low position close to the ground, which requires specific camera positions and angles. Therefore, occlusion will be a severe and unavoidable problem for datasets. For obtaining high-quality datasets in greenhouse environments, one method is to shoot manually, which is costly with a quantity of images. An alternative method is to obtain images automatically by monitoring cameras installed in greenhouses. Images in large quantities can be obtained at any time, which can meet the data demands of complex models such as deep learning models. However, since the spatial constraints, the location of fixed cameras is critical for obtaining high-quality images. Improper locations may cause serious occlusion of crops in the photos, which makes it difficult to identify targets. A feasible solution is to set up enough cameras to capture images in different growth periods of the crops. But a balance is needed between economic spending and image quality.

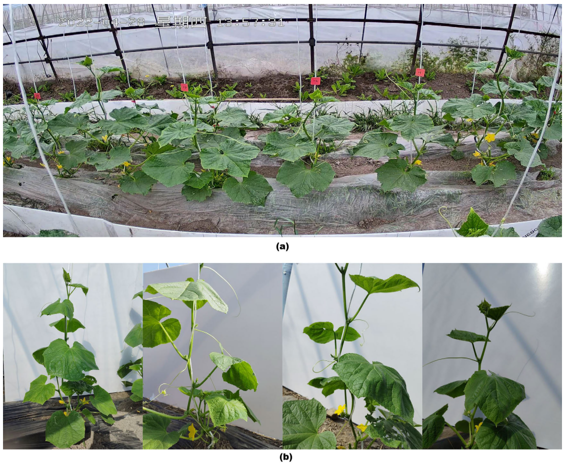
On the other hand, suitable models for cucumber flower detection in real-world greenhouse environments are still deficient. Detection models used for field plants or crops are hardly suitable for greenhouse vegetables due to their different characteristics. During cucumber cultivation, plants in the flowering stage grow with exuberant foliage and a high density, but the size of flowers is rather small compared with that of leaves. With existing technologies, it has been more difficult to detect small objects, and the detection rate is also lower than that of larger objects [21]. Images taken in in-field scenarios always have a complex and noisy background, which incurs an imbalance between the focused objects (foreground) and the background. Besides, occlusion is another challenging task. All these problems prevent the models from achieving a high detection accuracy.

In recent years, researchers have proposed a great number of methods for dealing with the problem of foreground-background imbalance and small object detection in the field of image processing [22], [23], [24], [25], [26], [27]. Data augmentation is one of the useful methods that has been used to increase the number of small object samples. Regular augmentation methods include rotation, translation and scaling of images. Some other technologies such as copy-pasting strategies and oversampling are also used for augmentation [28]. To alleviate the foreground-background imbalance, researchers have proposed new loss functions to improve the feature quality of images containing small objects [24], [29], [30]. Incorporating contextual information of small objects in images is another method to improve detection accuracy [31], [32]. However, it is still a challenge for technology communities to promote the low average precision of small object detection compared with large object detection.

In order to improve the final yield and conduct flower thinning, we attempt to detect cucumber flowers in this research to timely detect the number of flowers and identify flowers that are not growing well with computer vision technologies. In order to overcome the problems of occlusion and small target recognition, we have carried out the following work in this research: 1) Two cucumber flower datasets are built based on real-world greenhouse environments, which provide images taken by both monitoring cameras and hands. 2) Online and offline data augmentation methods are used for improving the sample number. 3) YOLO-based models with different attention mechanisms are proposed and compared in the detection of cucumber flowers. The main contributions of this paper are as follows:

(1) Datasets with cucumber flowers are provided, which enrich the current datasets of cucumber phenotypes in greenhouses.

(2) A newly designed model YOLOv5s-SE7 was used in the greenhouse environment to detect and classify cucumber


(a)

(b)

FIGURE 1. Image samples in the two datasets. (a) An image captured by camera with a complex background. (b) Images captured through manual photography with a white background.

flowers in different stages with an improved detection accuracy than reference models.

II. DATA

In this research, we investigated several public datasets for object detection, such as the famous Pascal VOC and MS COCO. Although these datasets contain plenty of categories of things including plants and flowers, there are few images of cucumber flowers in greenhouse environments. Hence, it is necessary to establish a novel dataset of cucumber flowers from scratch for the detection task.

A. DATASET

The images used in this work were taken under natural illumination conditions in April and May, 2022 in a greenhouse of the vegetable base in Yangzhou University ($32.29°N$, $119.48°E$), Jiangsu Province, China. Two image datasets were established for this research. One was collected by monitoring cameras (TL-IPC642-A4), which were set upon a movable support to simulate the effect at different locations. After adjusting the locations of the cameras, five or six plants could be shown in each image. Due to the actual production environment, all images in this dataset contained complex backgrounds, which contained greenhouse films and slings, as well as other vegetables and weeds, etc. The dataset included 100 JPG pictures of 30 cucumber plants in the field, the resolution of each of which was 2560*1440 pixels (Figure 1(a)).

Another dataset was collected through manual photography in September, 2022 in the same greenhouse. To shield the noisy background of the foreground plants, whiteboards were placed behind them, so that each crop could be more clearly presented in the images. Cameras of smart phones were fixed on a movable support to simulate the eyes of an observation robot. The whiteboards were moved manually to cover the complex background. The dataset included 380 JPG pictures with one plant in each. The resolution of each picture was 2560*1506 pixels (Figure 1(b)).

The images were labeled manually by bounding and tagging the cucumber flowers with an open-source tool named MAKESENSE (https://www.makesense.ai/), which helped to edit the labels across various platforms online. Because of the severe occlusion of cucumber leaves, stems and slings, a considerable proportion of flowers were only partially visible. Therefore, flowers in the pictures were labeled according to visible parts, which were annotated as much as possible in a way that people could identify.

B. DATA AUGMENTATION

The data volume of the first dataset was rather small for a deep learning (DL) network. To improve the performance of the DL model, data augmentation was used in this research to increase the diversity of samples. Several augmentation technologies were used, such as rotation, resize, shift, blur and noise, etc. After an offline augmentation, the number of images in the dataset increased to 1200, all of which were labeled by tools.

III. METHODS

A. DETECTION MODEL

In this study, cucumber flower detection depends on a deep learning model based on YOLOv5 (https://github.com/ultralytics/YOLOv5), which is one of the most widely-used algorithms in the YOLO family [33], [34], [35], [36], [37], [38], [39]. Different from the two stage models like the R-CNN family, YOLO is an end-to-end model with one stage for classification and regression [40], [41], [42]. Having advantages including a high speed and a small size, it is suitable to deploy YOLO on a mobile end [43], [44]. This merit makes it applicable for its operation in the field, which can be equipped on an intelligent agricultural robot such as an observation robot or a picking robot, so as to work in greenhouses.

B. MODEL ARCHITECTURE

YOLOv5 includes four different versions: YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, the difference among which is the number of residual components and convolution kernels. In order to use a light model applicable for mobile terminals in the greenhouse, we selected YOLOv5s as a baseline model (Figure 2), finetuned it and added attention models to improve the performance of our datasets.

The architecture of our improved model added an attention mechanism to the backbone network of YOLOv5s, so as to pay more attention to the features of the targets. The whole network consists of four parts: an input layer, a backbone, a neck and a head. The input images are adjusted to a specified size of 640*640 pixels in the input layer and the initial anchor size is set, which supports online augmentation technologies such as Mosaic. The backbone network includes a new CSP-Darknet network and SPPF to obtain the features of different depths within the images. There are four C3 blocks in the backbone, which have a similar effect to standard CSP

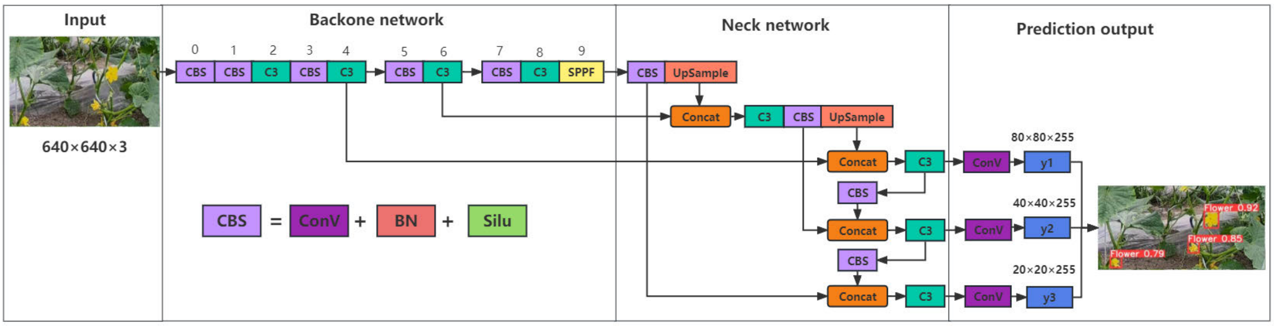


FIGURE 2. Model architecture.

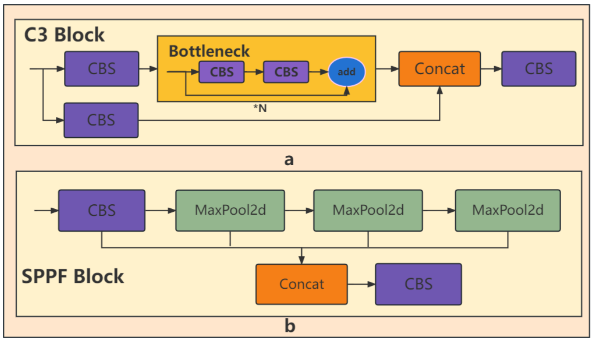


FIGURE 3. (a) C3 Block. (b) SPPF Block.

blocks. A C3 block includes three standard CBS layers and several bottleneck blocks to simplify the network structure and reduce the computation as well as reasoning time of models (Figure 3(a)).

The SPPF (Spatial Pyramid Pooling - Fast) layer is an improved version of the spatial pyramid pooling (SPP) network [45], [46], which improves the speed of models by changing the convolution kernel size of max-pooling to 5*5 and adjusting the concatenation of three max-pooling blocks (Figure 3(b)). The neck network includes an FPN-PAN structure to collect more complex semantic and location features [47], [48]. The head layer is used to predict the classes and locations of objects with feature maps of three sizes: 80*80, 40*40 and 20*20 pixels.

C. ATTENTION MODEL

An attention model is derived from the study on the selective characteristics of human vision, which has now been enormously popular as a component of neural networks leading to significant gains in the performance of a large number of applications [49]. More attention can be paid to the feature channels relevant to the objects while ignoring other irrelevant information. Network parameters will be updated by giving different attention weights for features in the propagation process

In order to overcome the problem of low detection accuracy of small objects with severe occlusion and a complex background in production environments, four attention models: SE, CA, CBAM and SimAM are adopted and compared in this research. SE means squeeze and excitation networks, which focus on the importance of different channels [50], [51]. CA refers to coordinate attention, which enhances the features in the width and height of an image respectively [52]. CBAM refers to a convolutional block attention model with two independent sub modules, namely the channel attention module (CAM) and the spatial attention module (SAM) [53]. CBAM focuses on not only channels but also spatial features. SimAM is another simple attention model of solutions with an energy function, which indicates the importance of a neuron in convolutional neural networks [54]. The attention models were deployed in a backbone network to improve feature extraction.

D. LOSS FUNCTION

The loss function is used to measure the difference between the predicted and real value of the model. The result of the loss function can be back-propagated and then the model parameters are updated. In YOLOv5s, the loss function is composed in three parts: the classification loss, the localization loss and the confidence loss.

The binary cross-entropy between the category label of the anchor box and the real category label is calculated based on the classification loss. The loss function of classification can be expressed as:

$$L_{cls} = -\sum_{i=1}^{N} (y_i \ln(\sigma_i) + (1 - y_i) \ln(1 - \sigma_i)) \tag{1}$$

$$\sigma_i = sigmoid(P(y_i = 1|x_i)) \tag{2}$$

where x_i and y_i denote the predicted category label and the real label respectively. N is the number of samples.

Localization loss reflects the offset of the predicted bounding box position from that of the ground truth bounding box, which is calculated through the complete-IoU (CIoU) method [55]. Through CIoU loss, the generalized IoU (GIoU) loss and distance-IoU (DIoU) loss are improved, both of which are used in the previous YOLO versions and good precisions have been achieved in YOLO as well as other visual processing models, such as Faster-RCNN [56].

Confidence loss is the loss function of object score of the model. The higher the score is, the greater the possibility for an object to appear in the bounding box will be. The loss of confidence is computed based on the same binary cross-entropy function as that of the classification loss.

The total loss is the weighted sum of the three losses above, where the weight coefficients λ_1, λ_2, and λ_3 are hyperparameters of the model. Letting L_{cls}, L_{loc} and L_{obj} denote the loss of classification, localization and confidence respectively, the formular of the total loss is:

$$Loss = \lambda_1 L_{cls} + \lambda_2 L_{loc} + \lambda_3 L_{obj} \tag{3}$$

When there is only one class to be detected, and the first term is zero in the total loss function.

E. MODEL TRAINING

In the task of object detection, models like YOLO always have numerous parameters to be determined through training. Therefore, a commonly-used method named transfer learning is adopted, that is, pre-train the model with a big dataset and then finetune it and its parameters to adapt to specific applications on the basis of the saved model. In our research, we used a pre-trained YOLOv5-6.0 model involving 118287 images of 80 COCO dataset classes and then finetuned the architecture as well as its parameters to achieve the best performance in our application.

In order to detect the cucumber flowers in the natural status of the greenhouse, we tested the detection effect of our model under the complex background of the environment. We divided our first dataset into the training set, the validation set and the test set according to the ratio of 8:1:1. Then our model was trained and tested based on the test set. 90% of the images in the second dataset were used as the training set and 10% were used as the validation set. No image was used for the test set because the complex background of the plants was shielded by a whiteboard in this dataset, which was inappropriate for the detection task in a natural growth environment. The dataset is just used to improve the accuracy of our model and reduce the labeling work of the images.

Furthermore, three stages of cucumber flowers are classified to investigate the classification effect beyond the detection. The three stages are defined as bud, bloom and faded flower, which represent that blooming takes less than one quarter, one quarter is taken for full blooming and after full blooming respectively. Different blooming stages reflect different statuses of cucumber growth, which is helpful for corresponding cultivation and management.

F. METRICS

Precision, recall and AP are adopted as the performance metrics of the models. Precision is the correct ratio of the samples, which is predicted to be positive. Recall refers to the correct ratio of the true positive samples. AP is the precision averaged across all recall values between 0 and 1 [57], which can be calculated based on the area under the precisionc × recall curve (AUC-PR). When there is only one category for detection, the metric mAP commonly used for multi-category detections is degraded to AP [58]. The value of the three metrics is between 0 and 1. The higher the values are, the better the detection performance of the model is. Formulars of the metrics are as follows:

TABLE 1. Detection results of the two datasets.

Dataset		P	R	AP@.5
First N=100	Val1	0.686	0.68	0.738
	Test1	0.643	0.734	0.755
Second N=380	Val2	0.881	0.775	0.880
	Test2	0.829	0.86	0.914
	Test1	0.379	0.381	0.317
First +Second N=480	Val3	0.773	0.789	0.848
	Test1	0.764	0.815	0.853

Note: Val and Test denote the validation set and the test set. 1,2,3 denote the first, second and first+second dataset respectively. N is the number of image samples in the dataset.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$AP = \frac{1}{N} \sum_{i=0}^{N} (r_{i+1} - r_i) P(r_{i+1}) \tag{6}$$

$$P(r_{i+1}) = max_{\widetilde{r} \geq r_{i+1}} P(\widetilde{r}) \tag{7}$$

In Formular 6, r_i is the ith different recall value of the detected object sorted by confidence. N means the number of different recall values. Note that N is usually less than the number of detected objects. $P(r_{i+1})$ is the maximum precision value among the precision values of the detected objects, whose recall value is larger than r_{i+1}. Usually, a positive sample is determined by IoU threshold, which is the ratio of intersections and unions between the candidate bound and the ground truth bound. In this research, the IoU threshold is set as 0.50 and the AP value is recorded as AP@.5.

IV. RESULTS

A. COMPARISON OF THE DETECTION EFFECT ON TWO DATASETS

The first dataset includes 100 images and the second includes 380. We tested the detection effect on the two datasets separately and jointly. When testing separately, the two datasets were both divided into 8:1:1. When training with the two datasets jointly, the second dataset was divided into 9:1:0. As it can be seen in Table 1, the result of precision, recall and AP@.5 is between 0.31 and 0.92. Although the AP@.5 value on the second test set (Test 2) is as high as 0.914, the performance of the model on the first test set (Test 1) is very poor with an AP@.5 of less than 0.32, which indicates that cucumber flowers in natural growth state with complex background interference cannot be accurately detected using the second dataset alone. However, when adding the first and second dataset to training the model, the detection result of Test 1 is improved significantly with AP@.5 from 0.755 to 0.853.

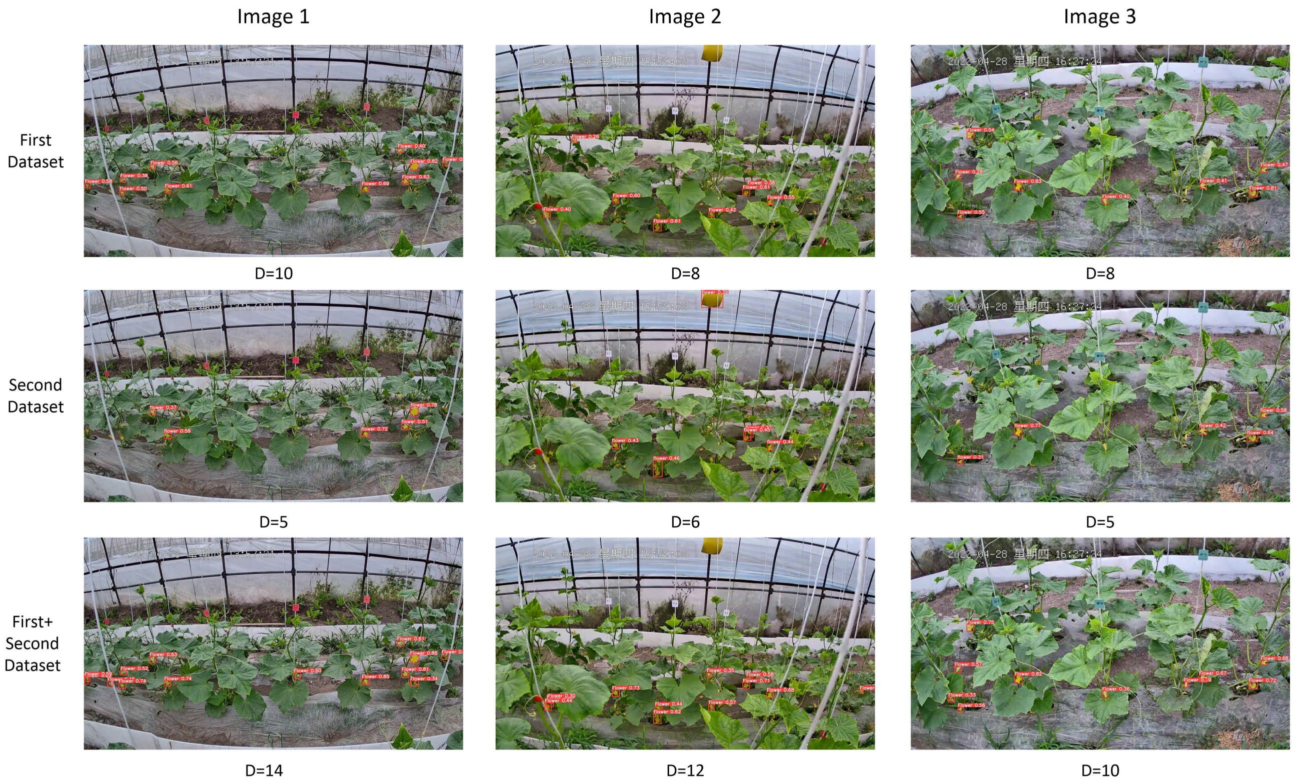


FIGURE 4. Comparison of detection results of the three images in the test set of the first dataset. The three images are displayed in three columns and D is the flower number detected in the images. On the left side of each row is the dataset corresponding to the model used by all images in the line.

Three images are presented intuitively in columns to display the detection results of the same test set (Test 1) with different models trained based on the datasets displayed in Table 1 (Figure 4).

The number of flowers detected in each image is provided below with letter D, which indicates that the model trained based on the second dataset (Line 2) has the worst detection effect with the smallest D in the column, while the model trained based on the first and second dataset jointly (Line 3) has the best effect. Some interference objects, such as the yellow sticky paper hanging above the plants in Image 2, are very similar to flowers in color and have caused misjudgment for the second model. Meanwhile through the model trained based on the first + second dataset, the number of flowers detected based on the first model with additional 4, 4 and 2 flowers are improved respectively in the three images, which reaches 40%, 50% and 25% in proportion. The results suggest that although the images in the second dataset have simple backgrounds with medium fields of view, they are effective in improving the recognition effect on complex scenes and large fields of view by increasing the diversity of samples.

B. EFFECT OF DATA AUGMENTATION

In this research, small datasets are used to reduce the work of manual labeling. Only 100 images containing cucumber flowers are selected in the first dataset, for which two augmentation methods are used, one is the online augmentation technologies provided by YOLOv5s, including mosaic, copy-paste and mix-up, etc., the other is offline augmentation applied manually. To investigate the effect of augmentation and provide clues for data preparations in similar research, different augmentation technologies are compared.

For online data augmentation, YOLOv5 provides several default options such as low-level augmentation, and we choose high-level data augmentation to achieve the best performance. For offline augmentation, two augmentation methods are compared: one is position-dependent, including: rotating 90°, 180° and 270° as well as flipping horizontally and vertically. The other is position-independent, including: blur, brighter, darker, scale and salt.

The results (Figure 5) of data augmentation show that high-level online augmentation provides obvious improvement in performance. The value of AP@.5 reaches 0.763 and 0.744 on the validation set and test set of online augmentation, which improves by 0.09 and 0.02 respectively compared with the original datasets with no augmentation.

Through location-dependent and -independent offline augmentation, the data volume is expanded by 5 times on the training, validation and test set individually. The AP value in Figure 6 indicates that location-dependent augmentation method is less effective than location-independent method,

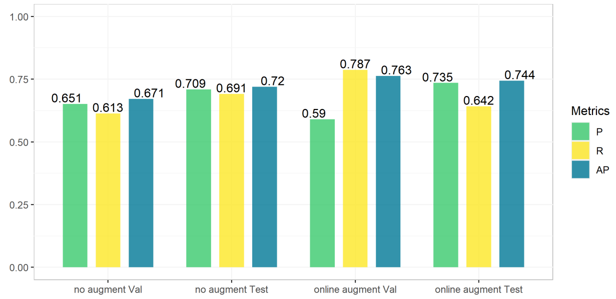


FIGURE 5. Comparison of online augmentation and no augmentation. Val and Test denote the validation set and the test set.

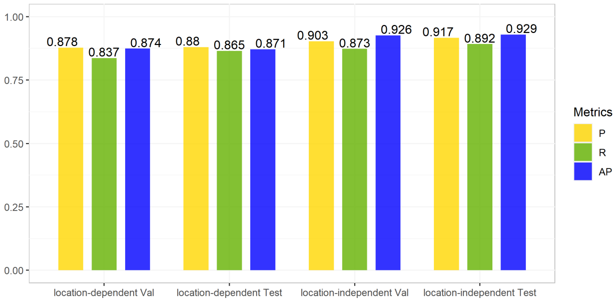


FIGURE 6. Comparison of location-dependent and -independent augmentation methods. Val and Test denote the validation set and the test set.

whose reason may lie in the characteristics of the images. For images captured in greenhouses, the generalization ability of the model may not be greatly improved due to position reversal or large angle rotation.

C. RESULTS OF ATTENTION MECHANISM

1) COMPARISON OF FOUR ATTENTION MODULES

To compare the effect of the four attention mechanisms added in YOLOv5s, both online and offline augmentation are used in the first dataset to achieve the best performance. Meanwhile the second dataset is used as the training and validation set. The total number of images reached 1580 with the first dataset expanded to 1200 (8:1:1) and the second dataset to 380 (9:1:0).

To focus the features of images in the backbone network and then transfer them to the neck of YOLOv5s, attention models are first deployed as Layer 9 in our improved model. The precision, recall and AP@.5 (Table 2) of the test set suggest that the SE attention mechanism deployed as Layer 9 has the best performance with an AP@.5 value that is 0.887 and 0.017 higher than that of baseline YOLOv5s model.

2) COMPARISON OF ATTENTION MODULES DEPLOYED ON DIFFERENT LAYERS

Through different layers of the backbone network of YOLOv5s, features of different depths in the images can be obtained. Adding attention models in different layers can help to focus on these features and get optimized results. Therefore, we investigated the effect of the four attention mechanisms placed on the early layers of Layer 7 and 5 of backbone network. As can be seen in Table 2, deployed as Layer 7, the SE model has the best effect among all attention mechanisms, then follow CBAM and CA. In addition, the effect of all the four attention mechanisms is better than that of baseline when deployed as Layer 7. However, the mechanism placed on Layer 5 is less effective.

TABLE 2. Performance comparison of the models with different attention mechanisms.

Model	P_{test}	R_{test}	AP@.5
Baseline	0.867	0.815	0.870
+SE9	0.859	0.831	0.887+0.017
+CA9	0.869	0.804	0.871+0.001
+CBAM9	0.834	0.818	0.876+0.006
+SimAM9	0.863	0.803	0.864-0.006
+SE7	0.855	0.863	0.905+0.035
+CA7	0.855	0.833	0.889+0.019
+CBAM7	0.885	0.825	0.900+0.030
+SimAM7	0.868	0.812	0.887+0.017
+SE5	0.850	0.710	0.836-0.034
+CA5	0.802	0.746	0.766-0.104
+CBAM5	0.837	0.755	0.808-0.062
+SimAM5	0.839	0.767	0.821-0.049

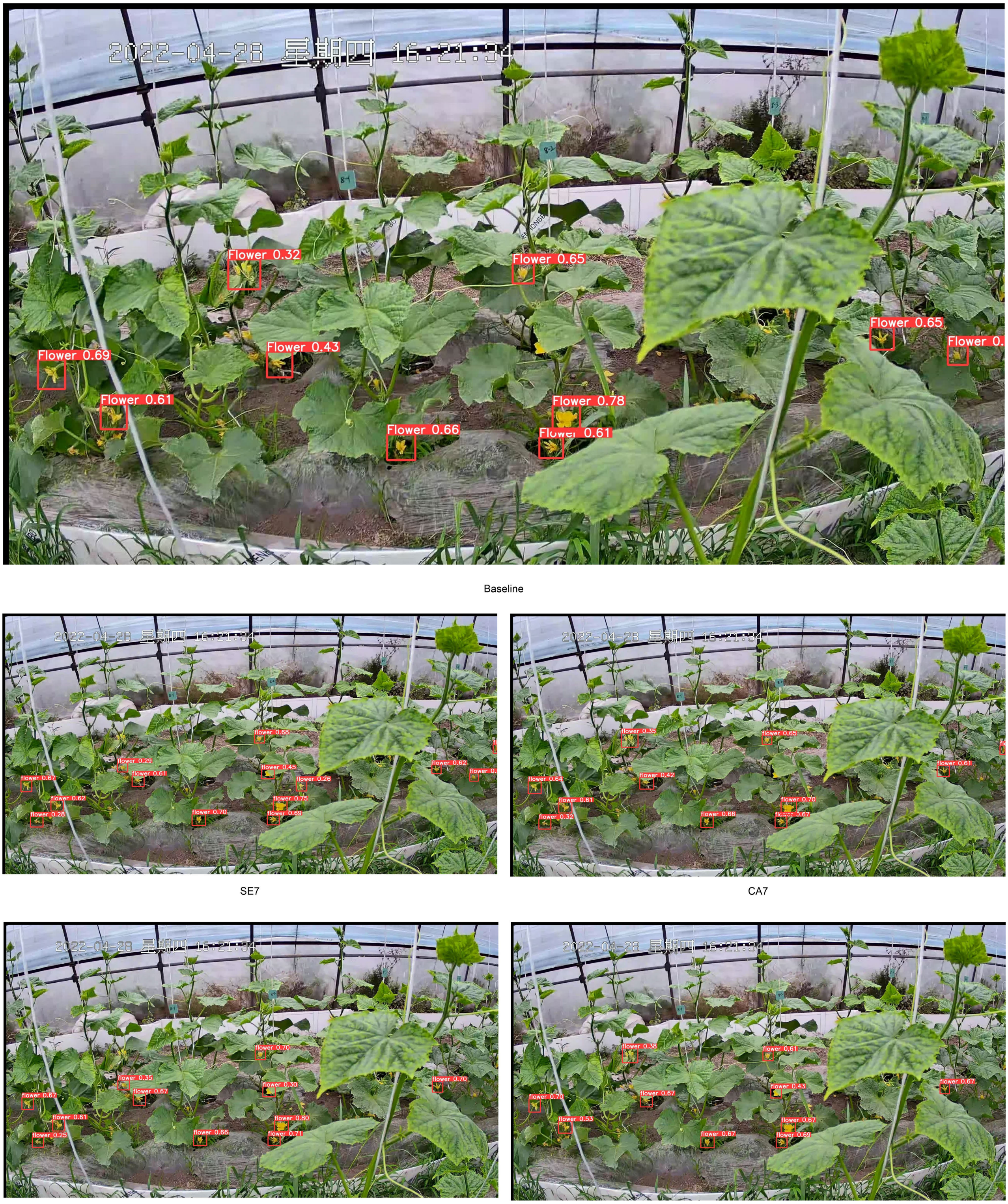
Note: The number after model name means the layer number in the backbone network.

The actual detected results of the test set (Figure 7) show that some flowers are not recognized by the models. For example, in Figure 8, a certain number of cucumber flowers on the left part of the image are not recognized because of their small size, and some others in the center and right part of the image are ignored by the models due to occlusion caused by leaves or stems. Meanwhile, the more complete the flowers are detected, the higher the confidence is. Among the four attention mechanisms, the best performance is achieved through the SE attention model, with 14 flowers detected. 11 flowers are detected through the CA and CBAM attention model. The same flowers are detected as the baseline YOLOv5s model with 10 flowers through SimAM attention models.

D. COMPARISON WITH OTHER MODELS

We compared the SE attention model (YOLOv5s-SE7) with other mainstream models in performance. The results in Table 3 indicate that YOLOv5s-SE7 model reaches the highest AP@.5 value among all the four models. Meanwhile, we investigated the average detection speed of each model with 10 images selected randomly from the test set, and the result showed that the baseline YOLOv5s model had the fastest speed, which was 0.01s for each image. As a two-stage model, the Faster-RCNN model was slightly slow, whose detection accuracy was lower than that of other models. The speed of YOLOv5s-SE7 model was next to that of the SSD model, but the accuracy of SSD was much worse than that of YOLOv5s and YOLOv5s-SE7.

Furthermore, the flower number observed by human and detected by YOLOv5s-SE7 model is fitted to investigate the



Baseline

SE7

CA7

CBAM7

SimAM7

FIGURE 7. The detection results of YOLOv5s based on different attention models deployed as Layer 7.

quantitative performance. The result presented in Figure 8 shows that 0.9 is achieved as the goodness of fit R^2, which indicates that the accuracy of the proposed model is close to that of manual detection.

E. DETECTION RESULTS OF CUCUMBER FLOWERS IN THREE STAGES

The three stages of cucumber flowers are identified to investigate the classification effect of the detection model.

TABLE 3. Comparison of the detection accuracy and speed of mainstream models.

Model	Model class	P	R	AP@.5	Time(second)
Faster-RCNN	Two-stage	0.733	0.646	0.666	0.0414
SSD	One-stage	0.819	0.475	0.709	0.0107
YOLOv5s	One-stage	0.867	0.815	0.870	0.0100
YOLOv5s-SE7	One-stage	0.855	0.863	0.905	0.0110

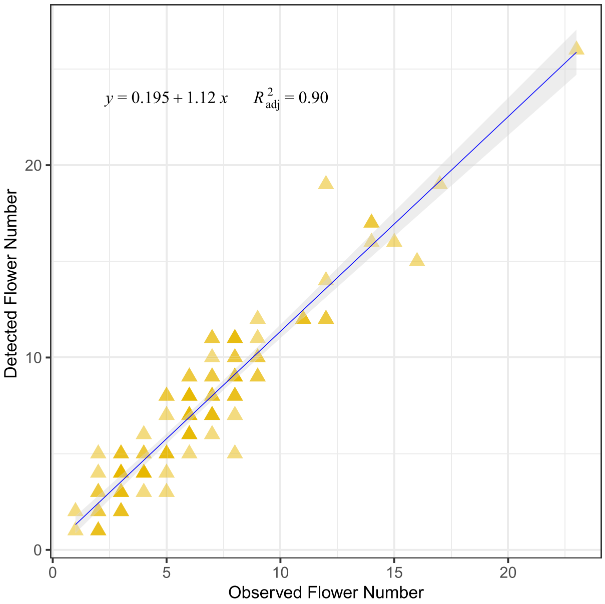


FIGURE 8. Comparison of flower numbers detected by human and YOLOv5s-SE7 model.

The classified results of 120 images in the test set are displayed in Table 4, which indicates that the detection performance in the bloom stage is better than that in the other two stages with AP@.5 of 0.92. The sample number in the faded flower stage is small and the accuracy is the lowest among the three classes. The mAP of all the three classes is approximately 0.85 with an IoU threshold of 0.5. Figure 9 is an example of detection results with three classifications.

V. DISCUSSION

A. DATASETS

Two datasets are used in this research. The first one is the primary dataset with a wide field of view, which reflects the actual growth environment captured by the cameras in the greenhouse. However, the images are full of complex backgrounds, making it difficult to identify the target flowers because of their small sizes. On the other hand, manual annotations in the images are also time-consuming and labor-intensive in the case of serious occlusion. Our previous experiments on COCO datasets show that the recognition effect will be reduced when labels are tagged incompletely (Figure 10). We selected 24919 from 30000 images in COCO datasets, each of which involves more than 10 labels. By deleting 10%, 30% and 50% of labels among them, we find that the average precision mAP of object detection is decreasing continuously from 0.409 to 0.182. The values of precision and recall also decrease significantly by 0.2 and 0.1 respectively. Therefore, when the first dataset is used alone, the annotations are likely to be incomplete because of the complex in-field environment, so the performance of the model is not very good.

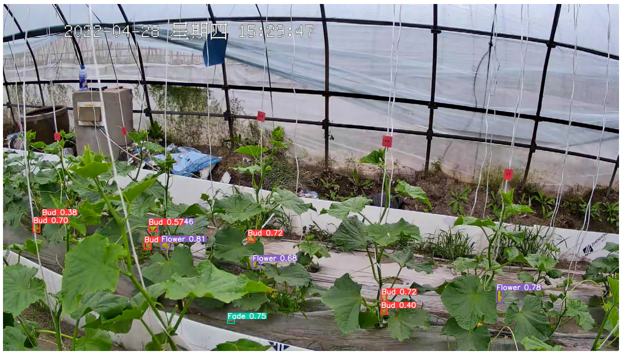


FIGURE 9. Example of detection results with three classifications. The red annotations represent buds, the purple ones represent blooms and the green ones represent faded flowers.

The second dataset is an auxiliary dataset used for model training, and images have been obtained with a medium field of view. As the background of plants is covered by a white board, it is easy for humans and models to detect flowers when labeling the images. Nevertheless, the actual scenes of vegetable growth are not so ideal. To create such photographic conditions, additional equipment such as a smart car with a white board on it could automatically move in the greenhouse to cooperate with the cameras. But it is difficult for planters to apply such equipment in ordinary plastic greenhouses due to extra costs and deployment difficulties. For example, the uneven ground in greenhouses requires smart cars to have a special caterpillar band for their ability to travel on this kind of ground. In addition, it requires certain techniques and skills to synchronize whiteboards with cameras. Therefore, we just take a small number of pictures manually to improve the recognition effect of models in terms of cucumber flowers in complex background environments. Nevertheless, the two datasets can be automatically collected in some sophisticated greenhouses.

B. DATA AUGMENTATION

The results of this study show that the accuracy of the model is effectively improved through data augmentation. Online augmentation is a key feature of YOLOv5. In this research, we tried high-level and low-level online augmentation, both of which demonstrated to be effective. But low-level augmentation is the default option for pre-trained models, therefore, it was more effective in our task. The AP@.5 value of the test set with low-level augmentation is 0.755 in the first dataset, which is 0.011 higher than that of high-level augmentation.

Different from online augmentation, offline data augmentation requires additional target annotations, which is time-consuming for manual labeling and quality inspection. Hence, it is necessary and labor saving for offline augmentation to label the targets automatically after data augmentation,

TABLE 4. The multi-classification results of YOLOv5s-SE7 detection.

Class	Labels	P	R	AP@.5	mAP@.5:.95
Bud	970	0.90	0.74	0.836	0.439
Bloom	651	0.90	0.87	0.921	0.605
Faded flower	137	0.80	0.67	0.783	0.413
All	1758	0.87	0.76	0.847	0.485

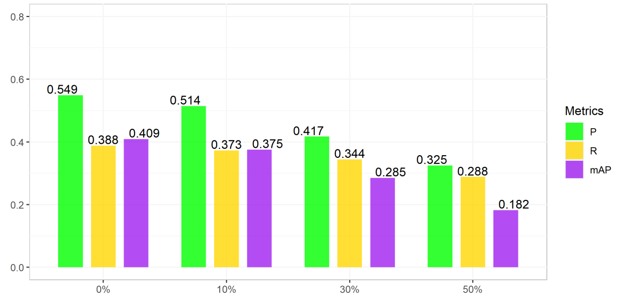


FIGURE 10. The detection effect of YOLOv5s on COCO dataset when the label incompletion is 10%, 30% and 50%. The dataset includes 24919 images selected from 30000 images, each of which contains more than 10 labels.

which is conducive to adding more images into the training set, so as to improve the accuracy of the models remarkably. On the other hand, the methods used for augmentation also need to be focused. By comparing the results of different augmentation, we found that the characteristics of the data needing to be considered when designing data augmentation methods and that more diversity features of the datasets should be provided in augmented data, which would help to improve the model performance.

C. ATTENTION MECHANISM

As attention mechanisms are deployed in the backbone network, different mechanisms and layers have different effect on the final detection results. The best performance is achieved by adding SE attention modules to Layer 7 of the backbone network in YOLOv5s. The reason may be that the color of cucumber flowers is very distinct from that of other objects such as green leaves in the images. Therefore, the SE attention, which has different importance in the feature maps of each channel, helps to extract the color features of flowers. In contrast, in CA and CBAM mechanism, more attention is paid to spatial features, which may make it more difficult to extract on account of the heavy occlusion and small size of flowers.

For three different layers deployed on the four attention mechanisms, the best performance appears in Layer 7. The reason may be related to the detection head of small, medium and large objects of the YOLOv5s model. Although the detection head of small objects is related to early layers of the backbone network, the comparison results in Table 2 show that the attention mechanism added to Layer 5 doesn't improve the performance obviously compared with that of the baseline. By observing the images in our datasets, we find that most flowers are less than 30×30 pixels within 640×640 images. Numerous literatures indicate that small objects ($<$ 32×32 pixels) or extremely small objects ($<$ 16×16 pixels) show a lower detection accuracy than larger objects [21]. The results of our study show an approximate performance in terms of cucumber flowers. The smaller the flowers are, the harder they are to be detected. But the difference is that in our task, only a very small part of many flowers can be shown due to the severe occlusion caused by leaves and other things in the greenhouse, which need some semantic information to increase the detection rate. Maybe it can explain why we can't simply extract shallow-level features to obtain good results. Stronger semantic information in the latter layers may also works.

VI. CONCLUSION

In order to intellectualize vegetable planting and reduce labor costs, automatic monitoring the growth and phenology of vegetables based on images becomes a promising method. In this work, we investigated the detection performance of YOLOv5s in terms of cucumber flowers in the environment of a real-world plastic greenhouse. Since there were no existing cucumber flower datasets for the task, we collected images and labeled flowers manually to create two datasets. Through online and offline data augmentation, attention mechanisms were added to the YOLOv5s network to improve the feature extraction of backbone network. The results showed that the SE attention module deployed as Layer 7 of the backbone network (YOLOv5s-SE7) presented the best performance when detecting small flowers from the noisy background of the greenhouse. This work can provide references for object recognition in agricultural production scenarios and support for intelligent counting as well as picking through smart robots in greenhouses.

The novelty of this research is that cucumber flower detection is a novel problem and the detection method under large field of view in greenhouse environments has not been reported before. This research aims to provide beneficial exploration by addressing the in-field recognition and classification of cucumber flowers. The images with medium field of view can effectively compensates for the issue of insufficient sample size of images with the large field of view, which provides a new approach for solving small sample problems by enriching data source for model construction. And the process of adding different attention mechanisms in different layers of the YOLO model to improve the recognition accuracy will provide valuable references for similar task when using attention mechanisms.

In summary, this work proposed a solution for small object recognition in agricultural production scenarios with complex background. In the future, the model can be used for intelligent flower counting and automatic flower thinning in greenhouses by smart robots.

REFERENCES

[1] S. K. Lonsbary, J. O'Sullivan, and C. J. Swanton, "Reduced tillage alternatives for machine-harvested cucumbers," *HortScience*, vol. 39, no. 5, pp. 991–995, Aug. 2004.

[2] Y. Zhou, L. Hu, J. Song, L. Jiang, and S. Liu, "Isolation and characterization of a MADS-box gene in cucumber (*Cucumis sativus* L.) that affects flowering time and leaf morphology in transgenic *arabidopsis*," *Biotechnol. Biotechnolog. Equip.*, vol. 33, no. 1, pp. 54–63, Jan. 2019.

[3] H. Zhao, X. Zhai, L. Guo, Y. Yang, J. Li, C. Ren, K. Wang, X. Liu, R. Zhan, and K. Wang, "Comparing protected cucumber and field cucumber production systems in China based on emergy analysis," *J. Cleaner Prod.*, vol. 236, Nov. 2019, Art. no. 117648.

[4] X. Wang, D. Gao, J. Sun, M. Liu, Y. Lun, J. Zheng, S. Wang, Q. Cui, X. Wang, and S. Huang, "An exon skipping in a *SEPALLATA-Like* gene is associated with perturbed floral and fruits development in cucumber," *J. Integrative Plant Biol.*, vol. 58, no. 9, pp. 766–771, Sep. 2016.

[5] D. Li, H. Zhao, X. Zhao, Q. Gao, and L. Xu, "Cucumber detection based on texture and color in greenhouse," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 8, Aug. 2017, Art. no. 1754016.

[6] Y. Sun, J. Zhang, H. Wang, L. Wang, and H. Li, "Identifying optimal water and nitrogen inputs for high efficiency and low environment impacts of a greenhouse summer cucumber with a model method," *Agricult. Water Manage.*, vol. 212, pp. 23–34, Feb. 2019.

[7] J. Chen, H. Qiang, J. Wu, G. Xu, Z. Wang, and X. Liu, "Extracting the navigation path of a tomato-cucumber greenhouse robot based on a median point Hough transform," *Comput. Electron. Agricult.*, vol. 174, Jul. 2020, Art. no. 105472.

[8] L. Xin-Xing, "Recognition method of cucumber leaves diseases based on visual spectrum and support vector machine," *Spectrosc. Spectral Anal.*, vol. 39, no. 7, pp. 2250–2256, 2019.

[9] X. Liu, D. Zhao, W. Jia, W. Ji, C. Ruan, and Y. Sun, "Cucumber fruits detection in greenhouses based on instance segmentation," *IEEE Access*, vol. 7, pp. 139635–139642, 2019.

[10] A. A. Jeny, M. S. Junayed, M. B. Islam, H. Imani, and A. F. M. S. Shah, "Machine vision-based expert system for automated cucumber diseases recognition and classification," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2021, pp. 1–6.

[11] J. Zhang, "Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things," *Int. J. Distrib. Sensor Netw.*, vol. 17, no. 4, 2021, Art. no. 15501477211007407.

[12] F. Wang, Y. Rao, Q. Luo, X. Jin, Z. Jiang, W. Zhang, and S. Li, "Practical cucumber leaf disease recognition using improved Swin transformer and small sample size," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107163.

[13] Z. Xiao, "Greenhouse intelligent recognition and replanting control system based on machine vision," in *Proc. ASABE Annu. Int. Meeting, Amer. Soc. Agricult. Biol. Eng.*, 2019, p. 1.

[14] W. Ji, J. Peng, B. Xu, and T. Zhang, "Real-time detection of underwater river crab based on multi-scale pyramid fusion image enhancement and MobileCenterNet model," *Comput. Electron. Agricult.*, vol. 204, Jan. 2023, Art. no. 107522.

[15] P. A. Dias, A. Tabb, and H. Medeiros, "Multispecies fruit flower detection using a refined semantic segmentation network," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3003–3010, Oct. 2018.

[16] Z. Cheng and F. Zhang, "Flower end-to-end detection based on YOLOv4 using a mobile device," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–9, Sep. 2020.

[17] J. Ye, M. Wu, W. Qiu, J. Yang, and W. Lan, "Polyphyletic loss: Litchi flower detection with occlusion," *J. Phys., Conf. Ser.*, vol. 2171, no. 1, Jan. 2022, Art. no. 012041.

[18] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Instance segmentation of apple flowers using the improved mask R–CNN model," *Biosyst. Eng.*, vol. 193, pp. 264–278, May 2020.

[19] J. Ärje, "Automatic flower detection and classification system using a light-weight convolutional neural network," in *Proc. EUSIPCO Workshop Signal Process., Comput. Vis. Deep Learn. Auton. Syst.*, 2019, pp. 1–5.

[20] L. Liu, H. Lu, Y. Li, and Z. Cao, "High-throughput Rice density estimation from transplantation to tillering stages using deep networks," *Plant Phenomics*, vol. 2020, pp. 1–14, Jan. 2020.

[21] G. Chen, H. Wang, K. Chen, Z. Li, Z. Song, Y. Liu, W. Chen, and A. Knoll, "A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 2, pp. 936–953, Feb. 2022.

[22] B. Bosquet, M. Mucientes, and V. M. Brea, "STDNet: Exploiting high resolution feature maps for small object detection," *Eng. Appl. Artif. Intell.*, vol. 91, May 2020, Art. no. 103615.

[23] Z. Liu, G. Gao, L. Sun, and L. Fang, "IPG-Net: Image pyramid guidance network for small object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 4422–4430.

[24] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Syst. Appl.*, vol. 172, Jun. 2021, Art. no. 114602.

[25] B. Bosquet, M. Mucientes, and V. M. Brea, "STDNet-ST: Spatio-temporal ConvNet for small object detection," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107929.

[26] B. Xu, X. Cui, W. Ji, H. Yuan, and J. Wang, "Apple grading method design and implementation for automatic grader based on improved YOLOv5," *Agriculture*, vol. 13, no. 1, p. 124, Jan. 2023.

[27] W. Ji, Y. Pan, B. Xu, and J. Wang, "A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX," *Agriculture*, vol. 12, no. 6, p. 856, Jun. 2022.

[28] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," 2019, *arXiv:1902.07296*.

[29] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[30] C. Deng, M. Wang, L. Liu, Y. Liu, and Y. Jiang, "Extended feature pyramid network for small object detection," *IEEE Trans. Multimedia*, vol. 24, pp. 1968–1979, 2022.

[31] C. Sun, Y. Ai, S. Wang, and W. Zhang, "Mask-guided SSD for small-object detection," *Int. J. Speech Technol.*, vol. 51, no. 6, pp. 3311–3322, Jun. 2021.

[32] J. Leng, Y. Ren, W. Jiang, X. Sun, and Y. Wang, "Realize your surroundings: Exploiting context information for small object detection," *Neurocomputing*, vol. 433, pp. 287–299, Apr. 2021.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[34] C. Liu, Y. Tao, J. Liang, K. Li, and Y. Chen, "Object detection based on YOLO network," in *Proc. IEEE 4th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Dec. 2018, pp. 799–803.

[35] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[36] H. Niu, X. Hu, and H. Li, "Improved YOLOv5 network-based object detection for anti-intrusion of gantry crane," in *Proc. 2nd Int. Conf. Control, Robot. Intell. Syst.*, Aug. 2021, pp. 147–152.

[37] H. Fu, G. Song, and Y. Wang, "Improved YOLOv4 marine target detection combined with CBAM," *Symmetry*, vol. 13, no. 4, p. 623, Apr. 2021.

[38] C. Xianbao, "An improved small object detection method based on YOLO V3," *Pattern Anal. Appl.*, vol. 24, pp. 1347–1355, Aug. 2021.

[39] Y. Huang, H. Cui, J. Ma, and Y. Hao, "Research on an aerial object detection algorithm based on improved YOLOv5," in *Proc. 3rd Int. Conf. Comput. Vis., Image Deep Learn. Int. Conf. Comput. Eng. Appl. (CVIDL ICCEA)*, May 2022, pp. 396–400.

[40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.

[41] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[42] M. L. Mekhalfi, C. Nicolò, Y. Bazi, M. M. A. Rahhal, N. A. Alsharif, and E. A. Maghayreh, "Contrasting YOLOv5, transformer, and EfficientDet detectors for crop circle detection in desert," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[43] M. Zha, W. Qian, W. Yi, and J. Hua, "A lightweight YOLOv4-based forestry pest detection method using coordinate attention and feature fusion," *Entropy*, vol. 23, no. 12, p. 1587, Nov. 2021.

[44] T. Jiang, C. Li, M. Yang, and Z. Wang, "An improved YOLOv5s algorithm for object detection with an attention mechanism," *Electronics*, vol. 11, no. 16, p. 2494, Aug. 2022.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[46] A. M. Basbug and M. Sert, "Acoustic scene classification using spatial pyramid pooling with convolutional neural networks," in *Proc. IEEE 13th Int. Conf. Semantic Comput. (ICSC)*, Jan. 2019, pp. 128–131.

[47] W. Zhan, C. Sun, M. Wang, J. She, Y. Zhang, Z. Zhang, and Y. Sun, "An improved YOLOv5 real-time detection method for small objects captured by UAV," *Soft Comput.*, vol. 26, no. 1, pp. 361–373, Jan. 2022.

[48] C. Dewi, R.-C. Chen, X. Jiang, and H. Yu, "Deep convolutional neural network for enhancing traffic sign recognition developed on YOLO v4," *Multimedia Tools Appl.*, vol. 81, no. 26, pp. 37821–37845, Nov. 2022.

[49] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 5, pp. 1–32, Oct. 2021.

[50] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[51] X. Zhong, O. Gong, W. Huang, L. Li, and H. Xia, "Squeeze-and-excitation wide residual networks in image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 395–399.

[52] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.

[53] S. Woo, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[54] L. Yang, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11863–11874.

[55] Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, and W. Zuo, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, Aug. 2022.

[56] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34. New York, NY, USA, Feb. 2020, pp. 12993–13000.

[57] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.

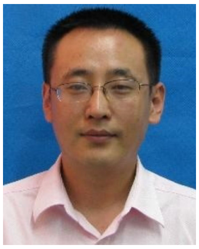
[58] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

XIANGYING XU received the B.S. and M.S. degrees in computer application from the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China, in 2001 and 2004, respectively, and the Ph.D. degree in agriculture from the Agricultural College, Yangzhou University, Yangzhou, China, in 2019.

She is currently a member of the College of Information Engineering, Yangzhou University. Her research interests include big data analysis, machine vision, and knowledge graph in agriculture.

HONGJIANG WANG was born in Wuwei, Gansu, China, in 1998. He received the B.S. degree from Yangzhou University, in 2021, where he is currently pursuing the master's degree.

His main research interests include artificial intelligence and object detection.

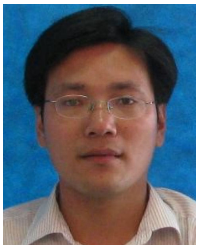


MINMIN MIAO was born in Rudong, Jiangsu, China, in 1973. He received the M.S. degree in agriculture from Yangzhou University, in 1996, and the Ph.D. degree in agriculture from Nanjing Agricultural University, in 2000.

Since 2019, he has been the Director of the Horticultural Department and a Ph.D. Supervisor with Yangzhou University. He is currently a Professor with the School of Horticulture and Plant Protection, Yangzhou University. He has presided seven national level scientific research projects. He has received awards and honors, such as the Third Prize for Science and Technology Progress in Jiangsu Province and the Outstanding Science and Technology Commissioner of Jiangsu Province.

WEIJIAN ZHANG was born in Baoji, Shanxi, China, in 2000. He received the B.S. degree from Xi'an Aviation University, in 2021. He is currently pursuing the master's degree with Yangzhou University.

His main research interests include artificial intelligence and object detection.



YONGLONG ZHANG received the Ph.D. degree in engineering from the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, in 2019.

He is currently with the Information Engineering College, Yangzhou University. He has presided multiple provincial and ministerial level projects. He has published more than ten academic papers in international journals or conferences. His research interests include cloud computing, mechanism design, and auctions.

HAIBO DAI received the B.S. degree from the Changshu Institute of Technology, Jiangsu, China, in 2016, and the M.S. degree from Yangzhou University, Jiangsu, in 2019, where he is currently pursuing the Ph.D. degree.

His research is mainly focused on how cucumber balancing the assimilate partitioning between its source and sink to get optimal yield.

ZIJIAN ZHENG was born in Nanjing, Jiangsu, China, in 1999. He received the B.S. degree from Yangzhou University, in 2021, where he is currently pursuing the master's degree.

His main research focus is on alleviating continuous cropping barrier in facility vegetable cultivation.

XIAOXIANG ZHANG was born in Dongtai, Jiangsu, China, in November 1980.

He is currently an Associate Researcher with the Lixiahe Agricultural Research Institute of Jiangsu Province, Yangzhou, China. He presided or participated in more than 20 provincial-level or national-level research projects. His research interests include rice cultivation and the development of physicochemical products with a focus on the absorption, transformation, and efficient application of nitrogen. He received eight awards from various scientific and technological achievements.