MDPI

Article

Reasearch on Kiwi Fruit Flower Recognition for Efficient Pollination Based on an Improved YOLOv5 Algorithm

Haili Zhou [1,2], **Junlang Ou** [1], **Penghao Meng** [1], **Junhua Tong** [1,2,*], **Hongbao Ye** [3] **and Zhen Li** [1]

[1] School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China; zy077520@163.com (H.Z.)
[2] Key Laboratory of Transplanting Equipment and Technology of Zhejiang Province, Hangzhou 310018, China
[3] Institute of Agricultural Equipment, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China
[*] Correspondence: jhtong@zstu.edu.cn

Abstract: A close relationship has been observed between the growth and development of kiwi fruit and the pollination of the kiwi flower. Flower overlap, flower tilt, and other problems will affect this plant's pollination success rate. A pollination model based on YOLOv5 was developed to improve the pollination of kiwi flowers. The K-means++ clustering method was used to cluster the anchors closer to the target size, which improved the speed of the algorithm. A convolutional block module attention mechanism was incorporated to improve the extraction accuracy with respect to kiwi flower features and effectively reduce the missed detection and error rates. The optimization of the detection function improves the recognition of flower overlap and the accuracy of flower tilt angle calculation and accurately determines flower coordinates, pollination point coordinates, and pollination angles. The experimental results show that the predicted value of the YOLOv5s model is 96.7% and that its recognition accuracy is the highest. Its mean average precision value is up to 89.1%, its F1 score ratio is 90.12%, and its memory requirements are the smallest (only 20 MB). The YOLOv5s model achieved the highest recognition accuracy as determined through a comparison experiment of the four sets of analysed models, thereby demonstrating its ability to facilitate the efficient target pollination of kiwi flowers.

Keywords: machine vision; target recognition; YOLOv5; kiwi fruit pollination

---

1. Introduction

Kiwi has a short flowering period, usually only 3 to 5 days, which increases the difficulty of pollination. Under natural conditions, kiwi pollination mainly relies on wind and insect pollination, but insect pollination is easily affected by weather factors. Female flowers cannot be fully pollinated, resulting in a low fruit-setting rate, small fruit, an increased deformed fruit rate, reduced economic benefits, and other problems, and a single kiwifruit needs at least 1000 seeds to produce more than 100 g of fruit [1].

Automatic targeting technology is rapidly developing. The automatic target-spraying machine studied by Jiang et al. [2] can selectively spray-pollinate a target in accordance with the changes in the target's position and characteristics, effectively improve the adhesion rate of the division on the crop, remarkably reduce the settlement of the division in the nontarget area, and obtain superior pollination performance, which can reduce production costs and effectively reduce farmers' burdens.

The key to automatic target pollination technology is target recognition and detection. Object recognition and detection mainly employ machine vision; accordingly, deep learning (DL) convolutional neural networks have developed rapidly in the field of machine vision in recent years and play pivotal roles in different fields. With the emergence of DL, more complex, in-depth models can now be constructed through continuous training with large sums of data so as to achieve automatic feature extraction and continuous optimization, thereby improving the accuracy of recognition. In 2012, the AlexNet network designed by

Krizhevsky et al. [3] won the ImageNet Contest by an absolute margin. Subsequently, more and better convolutional neural networks, such as VGG, ResNet, GoogleNet, YOLO and other networks, have been proposed.

Liu et al. [4] fine-tuned and modified YOLOv3 in terms of model feature weights and proposed the YOLOv3-SE model, which achieved better identification accuracy for Dongjujube in a dense environment where branches and leaves interweave and obscure targets. Zhang et al. [5] replaced the feature extraction network of YOLOv3 with the SEResNet50 network and proposed a new and improved model, ISDYOLOv3, which solved the problem of important information loss in the convolution process and achieved an average accuracy of 94.91% in the detection of mature mangos. Li et al. [6] realised the real-time monitoring of weeds in a cotton field by optimising the YOLOv3 model in the feature extraction network and building and testing the embedded platform. Based on the VGG network model, Yue et al. [7] added high-order residuals and parameter-sharing feedback subnetworks into the VGG network to identify crop diseases under complex natural conditions, resulting in higher identification accuracy (90.98%) and better robustness in actual environments.

Li et al. [8] proposed an efficient grape detection model, YOLO-GRAPE, which considered the complex growing environment of grapes. A subsampling fusion structure was added to the network to improve the accuracy of network recognition. Wang et al. [9] used the ResNeXt model with modified structural weight parameters to build a new feature enhancement module to address the lack of features in images with fog and used the attention mechanism to help the detection network focus on more useful features in such images. Lv et al. [10] proposed a BIFPN-S structure imitating the BiFPN model, which enhanced feature diffusion and feature reuse, and replaced the network SILU activation function with the ACON-C activation function to improve its network performance.

Chaschatzis et al. [11] used the DL method to detect the characteristic compressed tissue of sweet cherries. ERICA data sets, including an entire sweet cherry tree and single leaf images, were built to provide better recognition results over unbalanced data sets. Ficzere et al. [12] used the YOLOv5 algorithm to identify and classify the capsule defects of tablets and measure the capsule thickness, obtained a classification accuracy of 98.2%. Jin et al. [13] proposed a defect identification method based on DCGAN and YOLOv5 and solved problems regarding an insufficient number of samples and the uneven distribution of defect types in the defect detection of bonded structural parts by fine-tuning the structure and loss function of DCGAN. Li et al. [14] tested the food recognition algorithm of YOLOv5 on CFNet-34, a self-extended dataset based on the Chinese food dataset ChineseFood-Net. The average recognition accuracy of this algorithm was 89.7%, demonstrating good accuracy and robustness. Xue et al. [15] increased the attention mechanism and output layer to enhance feature extraction and feature fusion in order to improve recognition accuracy when analysing a complex background. The accuracy of target recognition was remarkably improved, and the mean average precision (mAP) value reached 80.5%. To detect postnatal defects of kiwifruit, Yao et al. [16] added a small target detection layer, improved the detection ability of the model for small defects, and introduced the loss function Ciou to improve the accuracy of regression. The CosineAnnealing algorithm was used to improve the effect of training. Zhao et al. [17] proposed a Wheat Grain Detection Network (WGNet) based on the training of the benchmark and used sparse network pruning and a mixed attention module to solve the problem of degradation. Zhang et al. [18] proposed an improved CoordConv feature extraction method to address the characteristics of weak echo intensity and small target area common in forward-facing sonar images, thereby endowing high-level features with corresponding coordinate information and improving the accuracy of network detection regression. Zhang et al. [19] proposed a method based on DL: YOLOv5-CA. A coordinated attention (CA) mechanism was integrated into YOLOv5 to highlight downy-mildew-related visual features and thus improve detection performance. Dai et al. [20] replaced Conv in the C3 module with CrossConv, which mitigated the problem of feature similarity loss in the fusion process and enhanced the module's feature

representation capacity. The SPP algorithm was improved by using the fast space pyramid pool algorithm, which reduces the number of feature fusion parameters and accelerates the feature fusion speed.

In the process of feature recognition, many researchers have proposed solutions to the problem of object overlap and occlusion. Gao et al. [21] integrated data enhancement to boost network generalizability. Moreover, non-maximum suppression (NMS) was optimized to improve the accuracy of the network, which improved its ability to recognize fruit overlap. Ye et al. [22] used methods such as data augmentation, Test Time Augmentation (TTA), and the Weighted Boxes Fusion (WBF) to improve the robustness and generalization of a model for the identification of terminal overlapping buds in different growth states. Li D H et al. ([23–25]) used the mean shift algorithm to pre-segment images and set the radius parameter range in the detection process to further accelerate the speed of the algorithm. Additionally, the centre coordinates and radius of the target were obtained by detection; then, the overlapping targets were identified.

In this study, we present the results of the use of an improved YOLOv5 neural network model to address certain problems, such as the inability to accurately identify the 3D position of kiwi flower buds in the natural environment due to the complex background images and lighting conditions present therein, which results in a large range of coarse pollination during pollination and the inability to accurately achieve the full pollination of flower buds. The target detection of kiwifruit flowers was conducted, the parameters were optimized, and the training parameters were adjusted in accordance with the shooting distance, the effect of capturing the pictures, and the intensity of the flowers so as to achieve the highest accuracy in terms of target detection. In this study, the method of single target and double frame was adopted. The angles of the flowers were accurately identified through the position relationship of the flower and the stamen selection box, combined with the actual angles of the flowers, which provided angle data to support the accurate detection of the target. The pollination strategy of overlapping flowers was studied. Based on the pollination strategy, the detection function of YOLOv5 was modified such that it would meet the ideal requirements. A double-flow nozzle was used to pollinate the flowers, the spray parameters were determined, and the pollination area relationship of the inclined flowers was analysed. The droplet escape rate was measured, the spray compensation time was calculated, and a verification test was performed to prove the feasibility and practicability of this study and provide reliable support for the precise selection of kiwifruit using target spray technology.

2. Materials and Methods

2.1. Image Acquisition

A PyTorch DL framework was built, and PyCharm platform was used to achieve the training and testing of the model with respect to kiwi flower and bud recognition on a desktop computer (Intel core i5 10400fCPU, 2.6 GHz, 16 G memory, NVIDIA GeForce RTX 2060GPU, 6 GB video memory, Windows10 system, and CUDA and Cudnn libraries) made by the Intel Corporation, which is located in Arizona, the United States. The initial learning rate was set to 0.01, the threshold of IOU was set to 0.01, 8 samples were used as a batch-processing unit during model training, and the number of training rounds was 200. The optimal training weight parameter file was obtained after model was trained. The optimal weight file was used to distinguish kiwi flower pictures, and the performance of the recognition model was evaluated in accordance with the identification scenario.

A red kiwi plantation in Qiaosi, Hangzhou, was selected as the base of this experiment, as shown in Figure 1. The size of the kiwifruit plantation greenhouse was 6 m × 50 m and 150–170 cm off the ground, with melon rack spacing of 1.5 m. The light conditions in the greenhouse were sufficient, the kiwifruits were planted orderly, and the flowers were oriented in different directions, wherein about 70% of them were oriented downward, while the rest were scattered. The angle data must be identified and obtained to provide data support for the precise pollination process of the target.
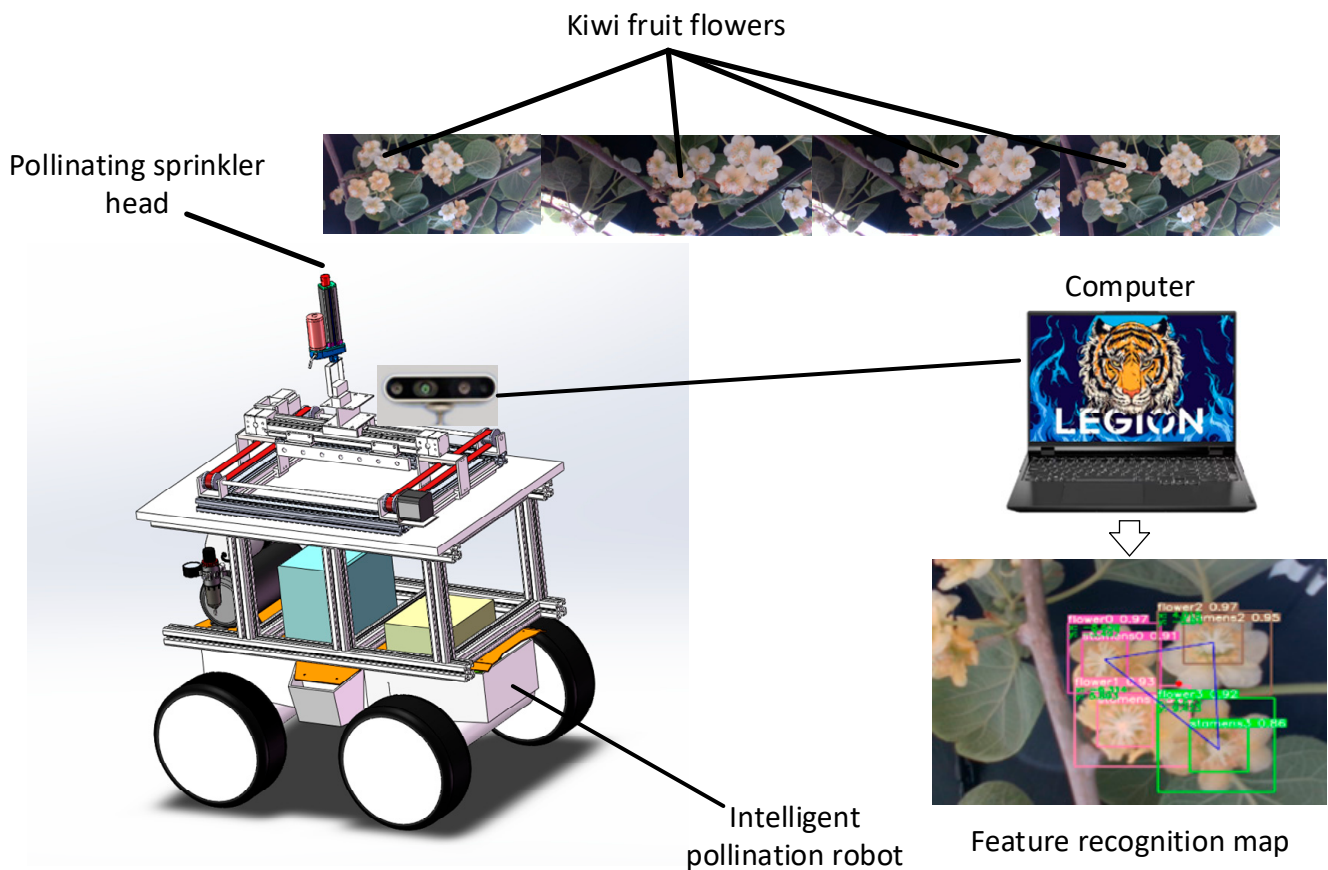
Figure 1. Kiwi flower feature recognition process.

An Intel RealSenseD415 camera made by the Intel Corporation, which is located in Arizona, the United States, was selected for this study. It is equipped with D400 series depth module, and its field-of-view angle is about 70°. It can effectively capture 10 m area within sight and supports the output of 1280 × 720 resolution depth. D415 has high pixel density and can achieve better, more accurate scanning for users in any given area of the same point. It has a special antiglare-processing function and can be used outdoors. The shooting height range of the camera in this study was 30–50 cm, and the camera was placed horizontally so that it faced upward when shooting.

2.2. Based on YOLOv5 Algorithm

There are a variety of object detection algorithms, such as YOLOv5, MobileNet, SSD, faster R-CNN, VGG, ResNet, etc. Compared with YOLOv5, MobilleNet, as a lightweight deep neural network, has fewer parameters and higher precision and can generate required feature maps with less computing time. However, due to its small number of convolutional layers, its feature extraction ability is very insufficient. The SSD model is fast but has poor recognition performance for small objects. Faster-RCNN uses RPN to generate candidate regions, and then extracts features through RoI pooling, so its accuracy is relatively high, and its processing speed is fast. However, its time complexity is high, it is difficult to use in practical applications, and its training and reasoning times are extensive. The VGG model has fewer parameters and operates through the convolutional series method, and it is stable and easy to transplant. However, due to its large number of full connection points and deep network structure, its training speed is slow. The residuals block is added to the ResNet network, which is helpful for the back propagation of the gradient in the training process, but there is a great deal of redundancy in the deep residuals network. In conclusion, the yolov5 model can significantly separate features, enhance mesh feature fusion, and ensure faster training speed.

YOLOv5 is a further improvement of the YOLOv4 algorithm, offering superior detection performance and rendering the results of testing via the coco test set more remarkable than before. The main innovation of YOLOv5 is its integration of multiple shortcuts. Accordingly, this algorithm can be used to develop a set of fast training and deployment schemes.

The network structure of YOLOv5 algorithm is mainly composed of input terminal, backbone network, neck network, and prediction component, and its network structure is shown in Figure 2. YOLOv5 can be divided into four models, namely, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, in accordance with different depths and feature widths of the models. The larger the model, the greater its detection accuracy. However, despite its longer detection time, the YOLOv5s model is the smallest and most suitable for rapid detection tasks, and it offers superior detection accuracy. It can meet the requirements of daily agricultural and industrial applications.
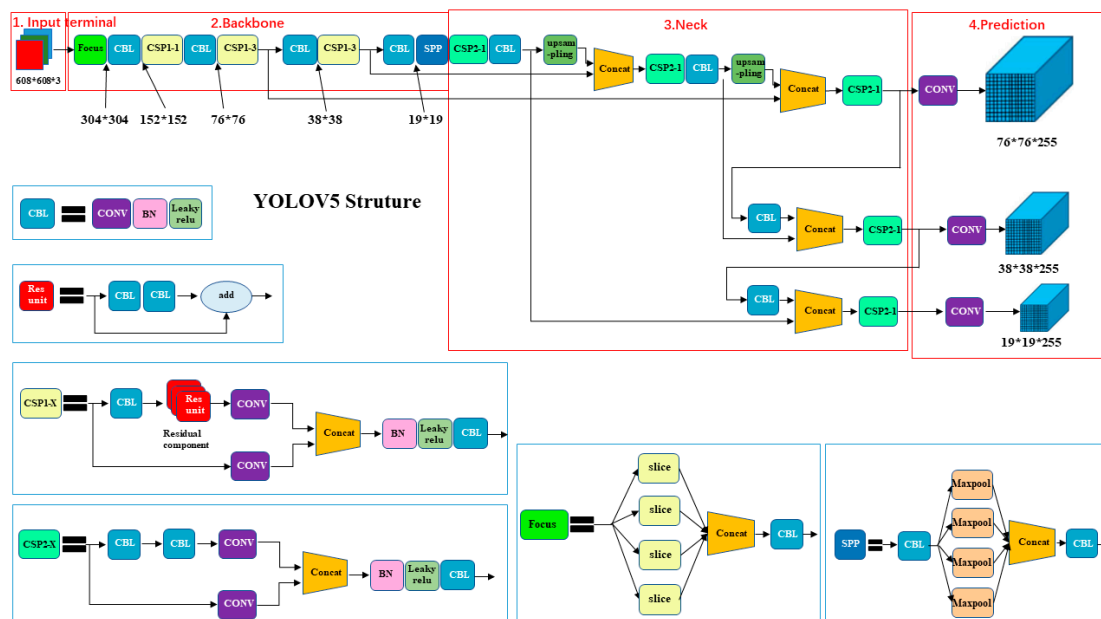

Figure 2. YOLOv5 frame diagram.

First, YOLOv5 was used to enhance the mosaic data of the input image at the input end and then randomly selected four photos in the data set to be pieced together into one by means of random scaling, cropping, and arrangement for training. This step enriched the data set, improved the training speed of the network, and reduced the memory requirements of the model. Adaptive anchor frame calculation and adaptive picture scaling were conducted together with stitching. In the network-training stage, the prediction box corresponds to the initial anchor frame output, and the size of the input picture is not limited. The pictures of different sizes are fitted to a fixed, common size; then, the pictures are input into the detection network. In this way, the image input process is simplified.

The Focus structure and SPP structure are added to the BackBone network. The focus structure can slice the input image. The original $608 \times 608 \times 3$ image outputs a feature mapping of $302 \times 302 \times 32$ after slicing, Concat, and Conv operations. The SSP module improves the receptive field, remarkably separates the features, and strengthens the network feature fusion ability so that images of any feature size can be inputted, and the size of the input images can be diversified.

The feature pyramid (FPN) + path aggregation structure (PAN) is added to the Neck network, which can allow for image feature fusion from multiple angles. Its complete structure is composed of an FPN layer that conveys semantic features from the top down and a PAN layer that conveys strong positioning features from the bottom up. Under

the synergistic action of the two layers, the feature extraction effect is enhanced, and the precision of network training is improved.

In the prediction phase, the object sizes of the extracted and spliced feature maps are predicted and classified, and the predicted results are compared with the actual measured results and optimised. As shown in the Figure 2, the model outputs three sizes of the prediction models, and the models of different sizes have remarkable differences in terms of the field of sensitivity and resolution. An image output with a small multiple is suitable for the detection of small objects, whereas an image output with a large multiple is suitable for the detection of large objects. The output multiple of the image should be selected in accordance with the size of the detected object.

2.3. Methods Used in This Study

The main working methods used in this study are as follows: The K-means++ clustering method was used to cluster out the prediction anchor frame closer to the target size, and convolutional block attention module (CBAM) mechanism was added. *CIOU* was used to replace *GIOU*. The modules for calculating the flower inclination angle, judging the overlap, determining the pollination point coordinates, and searching for the pollination angles were added into the detection function.

2.3.1. Use of K-Means++ to Cluster out New Anchor Boxes

By default, YOLOv5 uses the K-means algorithm to cluster data sets for generating anchor boxes and uses genetic algorithm to adjust the size of anchor boxes during training. The k-means clustering algorithm initialises K points randomly as clustering centres, calculates the distance between samples and each clustering centre, divides the samples to the nearest clustering centre point, calculates the mean value of all sample features divided into each category, and uses the mean value as the new clustering centre. However, the initial random selection of clustering centre is uncertain, which affects the convergence speed of the algorithm or leads to false classification. Therefore, this study uses the K-means++ clustering algorithm to obtain a prior anchor box that is more consistent with the data sample.

The essential difference between k-means++ and k-means algorithms depends on the initial selection of K clustering centres. The basic principle of K-means++ in the initialisation process of the cluster centre is to ensure that the distance between the initial cluster centres is as far as possible. First, a sample point is randomly selected as the initial cluster centre; then, the sample point with the largest distance is selected as the next cluster centre through the roulette wheel method until K cluster centre points are selected. Finally, the K-means algorithm is used to determine the final clustering centre. Although the K-means++ algorithm takes a certain amount of time for the initial determination of cluster centre points, it reduces the influence of the random selection of cluster centre points on the size of anchor frame selection, selects better initial cluster centres, and improves the speed of the algorithm.

2.3.2. Adding CBAM Mechanism

When observing objects, the human visual system will transfer more attention to areas whose objects' attributes can be quickly judged. This type of attention selection mechanism can accelerate our recognition of an object's attributes and use attention resources more effectively. In this study, the attention mechanism is generated by referring to this feature of the human eye. In essence, the model allocates a certain weight to different features of an object and draws attention to the most useful features of the object so as to improve the object recognition rate.

Attention is divided into two main types, namely, spatial attention (CAM) and channelled attention (SAM), as shown in the figure. Compared with other attention mechanisms, CBAM attention mechanism is a simple and effective attention module for feedforward convolutional neural networks, which consists of two modules, namely, spatial attention

module (CAM) and channel attention module (SAM). For the input feature map, the CBAM module inferences the attention diagram in two separate dimensions (channel and space). The CAM module first applies global maximisation and global average pooling to the input feature graph, transmits the processed feature graph to a shared neural network, performs elementwise addition operation on the output feature, and outputs the attention feature through calculation. The attention features generated by the CAM are input into SAM after multiplication. The spatial attention features are obtained through sigmoid nonlinear operation in SAM. Finally, the attention map is multiplied by the input feature map to perform adaptive feature optimisation, save parameters and computational power, and ensure that the function can be realised immediately in the existing network.

2.3.3. CIOU Replaces GIOU

By default, YOLOv5 adopts *GIOU*-loss as the bounding box's loss function and uses binary cross entropy and Logits loss functions to calculate the loss probability and target score. However, the loss function is the same as *IOU*-loss when the values of *IOU* and *GIOU* obtained during calculation are the same, which makes it impossible to measure the real position relationship. Therefore, in this study, *CIOU*-loss is chosen to replace *GIOU*-loss, and the formula is expressed as

$$CIOU = IOU - \frac{\rho^2\left(b, b^{gt}\right)}{c^2} - \alpha v \tag{1}$$

CIOU provided two penalty terms, and box regression considered the coverage area, centre distance, and aspect ratio, which more effectively solved the three problems and allowed the model to obtain a better regression effect. As shown in Figure 3.
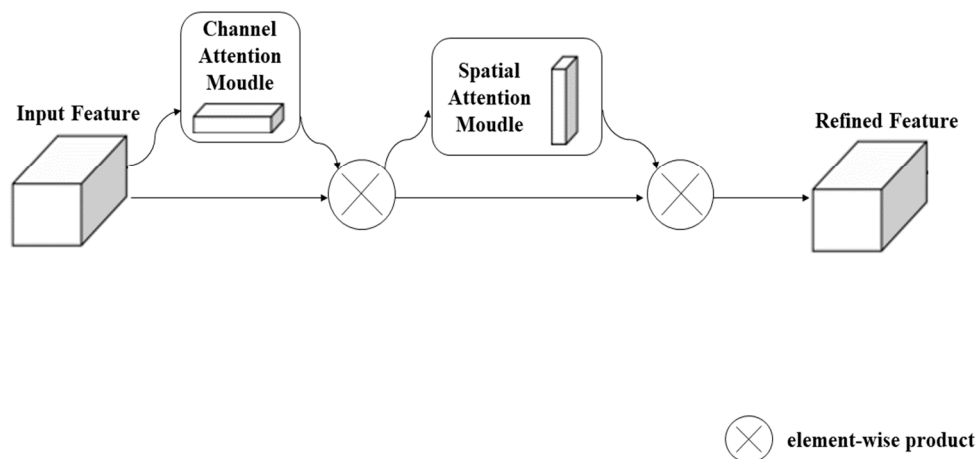
Convolutional Block Attention Module

Figure 3. Attention model.

2.3.4. Detection Function for Flower Angle Calculation Module

The research shows that different locations of the recognition box will have a distinct influence on the detection of the tilt angle of the flower. When the distance between the recognition box changes, the detected tilt angle of the flower will also change. When the distance between the centres of two flower frames is minimal, the flower tilt can't be detected, but when the distance between the centres of two flower frames is large, the detected flower tilt will show a certain error.

In accordance with the above research and analysis, a certain correlation is observed between the tilt angle of the flower and the position of the recognition box. Therefore,

we needed to establish a model including the flower and the flower core. We can further explore the position relationship between the flower tilt angle and different flower's centre distance through the calculation and analysis of this model. As shown in Figure 4, the original length of the flower is *L*, the horizontal length of the flower is *b*, the distance between the centre point in the horizontal direction of the flower and stamen is Δx, and the thickness between the flower and the stamen is a when the tilt angle of the flower is α.
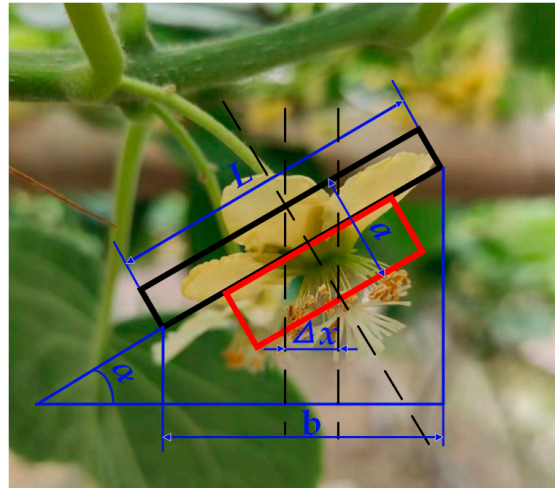
Figure 4. Flower tilt model.

As shown in Figure 4, the tilt angle of flowers has the following functional relationship with the horizontal length of flowers and the thickness of flowers and stamens

$$\cos\alpha = \frac{b}{L} \tag{2}$$

$$\sin\alpha = \frac{\Delta x}{a} \tag{3}$$

The formula for calculating the tilt angle of the flower can be obtained by combining the above two formulas and simplifying the calculation, as follows:

$$\alpha = \arctan\left(\frac{L}{a} \times \frac{\Delta x}{b}\right) \tag{4}$$

We can find the expression of the tilt angle of the flower by using this formula. However, the flower's own gravity has an effect on the actual tilt in a real scenario. Considering that the flower stalk is long and not rigid, it is prone to deformation and bending under the influence of the gravity of the flower itself after the flower tilts at a certain angle, thereby reducing the tilting angle. We conducted field research in an orchard and collected relevant pictures and data of flowers tilting at a certain angle while considering the influence of gravity on the tilt angle of flowers, as shown in Figure 5. By combining the actual data and the calculated tilt angle, the calculation formula of the tilt angle of the flower was further optimised and improved, and the accuracy of the calculation formula of the tilt angle of the flower was improved.

Figure 5. Flower tilt photo.

The collected pictures and data were processed by using the regression function to reduce the adverse influence of gravity on the calculation of the actual tilt angle of flowers, and the final regression formula can be obtained as follows:

$$y_a = 8.663\sqrt{\frac{100\Delta x}{b}} \tag{5}$$

where y_a is the calculated angle. This formula can satisfy the calculation of the tilt angle of flowers, and the error is controlled within 5%. In accordance with the actual situation, the maximum tilt angle of flowers is 30°.

The error between the calculated tilt angle and the actual measured tilt angle is calculated as follows

$$Y = \frac{|y_b - y_a|}{y_a} \times 100\% \tag{6}$$

where Y is the error rate and y_b is the actual tilt angle by experimental measurement.

2.3.5. Search for Pollination Points Based on Flower Overlap and Flower Angle Identification

The analysed flowers have four distribution conditions: single flower, two overlapping flowers, three overlapping flowers, and four or more overlapping flowers. In accordance with the spray pollination method, the final pollination area is a circle with a certain area. Therefore, a pollination strategy based on polygons was developed to find the edge points of polygons and thus ensure that the flowers can be fully pollinated within a given range. The intersection of two flower frames is defined as an overlap point. The specific pollination strategies are as follows.

For a single flower, the centre point of the flower core is selected as the pollination point; thus, the nozzle moves toward the pollination point, and the nozzle angle is not tilted.

When two flowers overlap, the centre point of the two flowers is selected as the feature point, and the two feature points are connected. The centre point of the feature point is designated as the position of the nozzle. The nozzle tilts at a certain angle after it moves to this position and is then aimed at the two flower core feature points for pollination.

When three flowers overlap, the centre point of the three flowers is selected as the feature point, and the three feature points are connected to form a triangle. The centre of the triangle is selected as the position of the nozzle. The nozzle tilts at a certain angle after it moves to this position and is aimed at the three flower core feature points for pollination.

When four or more flowers overlap, the top, rightmost, bottom, and leftmost feature points are searched. If four feature points are found, then the quadrilateral centre is selected as the position of the nozzle. The nozzle tilts at a certain angle after it moves to this position

and is aimed at the four flower core feature points for pollination, ensuring that the entire overlapping area is covered. Three feature points are regarded as three flowers overlapping, and two feature points are treated as two flowers overlapping. The specific pollination scenario is shown in Figure 6.
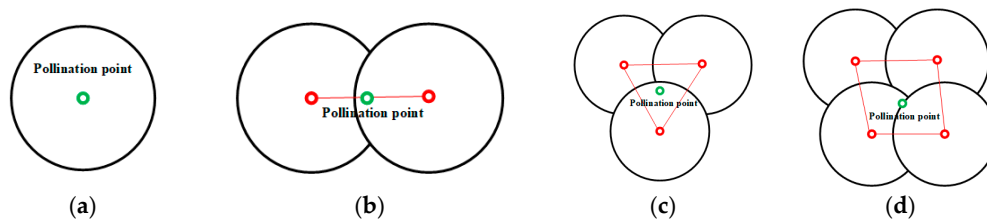

Figure 6. Schematic of pollination points under overlapping conditions: (**a**) single flower; (**b**) overlap of two; (**c**) overlap of three; (**d**) four or more overlapping.

In accordance with the overlap relationship, the recognition results of overlap is considered to be matched with the actual when the intersection area of two flower frames accounts for more than 60% of the flower frame area. In accordance with the Gauss area shoelace formula, the centroid of the polygon in the overlapping case is calculated. The area of the polygon is calculated in accordance with the shoelace formula.

$$A = \frac{1}{2} \left| \sum_{i=1}^{n} x_i(y_{i+1} - y_{i-1}) \right| = \frac{1}{2} \left| \sum_{i=1}^{n} y_i(x_{i+1} - x_{i-1}) \right| = \frac{1}{2} \left| \sum_{i=1}^{n} det \begin{pmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{pmatrix} \right| \quad (7)$$

where A is the polygon area of overlapping flowers, x_{i-1}, x_i, x_{i+1} is the abscissa of the overlapping flower centroid, y_{i-1}, y_i, y_{i+1} is the ordinate of the overlapping flower centroid. The position of the centroid can be calculated by using the root equation, as follows:

$$c_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (8)$$

$$c_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (9)$$

where c_x is the abscissa of the pollination point, c_y is the ordinate of pollination point.

The centroid's coordinates are outputted, and the tilt angle with respect to the target is calculated in accordance with the coordinates of the centre points of the stamens of each overlapping polygon, as shown in Figure 7.
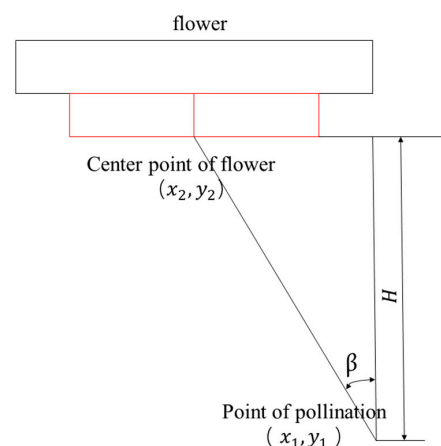

Figure 7. Pollination of the centre point of the flower.

The formula for calculating the pollination angle is as follows:

$$\beta = \frac{|x_1 - x_2|}{H} \tag{10}$$

where β is the pollination angle, x_1 and x_2 are the horizontal coordinates of the pollination point and the flower centre point, respectively, and H is the pollination distance.

In accordance with the above regression function of tilt angle and overlapping pollination strategy, the detection function can be modified, as shown in Figure 8, and the tilt angle calculation module, overlapping flower scenario identification module, and pollination point discovery module can be incorporated.
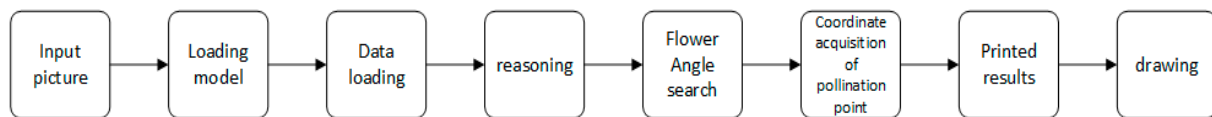

Figure 8. Image-reasoning process diagram.

The input image is identified in accordance with the weight obtained from the training data set. The setting of the double-target frame will identify the flowers and stamens and match them in accordance with their intersection relationship. The tilting angle of the flowers in the X and Y directions after matching will be determined by using the regression equation of the tilting angle of the flowers; then, the overlap situation of the flowers will be determined in accordance with the intersection ratio. Thus, the coordinates of pollination points and the angle at which the nozzle tilts with respect to the overlapping flowers in X and Y directions at the pollination points are determined in accordance with different overlapping conditions, and the results are printed and drawn.

2.4. Model Evaluation

The precision, recall rate, and mAP value of the model are used as the evaluation criteria for the feature recognition quality of YOLOv5 model, which can be calculated as follows:

$$Precision = TP/(TP + FP) \tag{11}$$

$$Recall = TP/(TP + FN) \tag{12}$$

$$F1 = [2/(1/Precision) + (1/Recall)] \tag{13}$$

$$AP = 1/11 \times \sum\nolimits_{r \in (0, 0.1, 0.2, \ldots\ldots 1)} pinterp(r) \tag{14}$$

$$mAP = \frac{\sum AP}{N} \tag{15}$$

where *Precision*—the ratio of true position in the recognized image; *Recall*—the proportion of the left and right positive samples in the test set that are correctly identified as positive samples; *F1*—the weighted harmonic average of Precision and Recall; *AP*—the interpolated average precision of the detection algorithm; *TP*—the result wherein kiwi flowers are correctly identified; *FP*—the result wherein kiwi flowers are incorrectly identified; and *FN*—the result wherein kiwi flowers are not detected.

2.5. Data Set Construction and Model Parameter

In the test base shown in Figure 9, more than 2000 kiwi flower pictures with a resolution of 1920×1080 were collected. The shooting period was from 3:00 pm to 5:00 pm, and multiple shooting angles were selected During the shooting of kiwifruit orchard, the distribution of kiwifruit flowers in full bloom was complex, and many overlapping situations were observed. As shown in Figure 10, the main overlapping situations consisted of two flowers overlapping, three flowers overlapping, and four or more flowers overlapping.
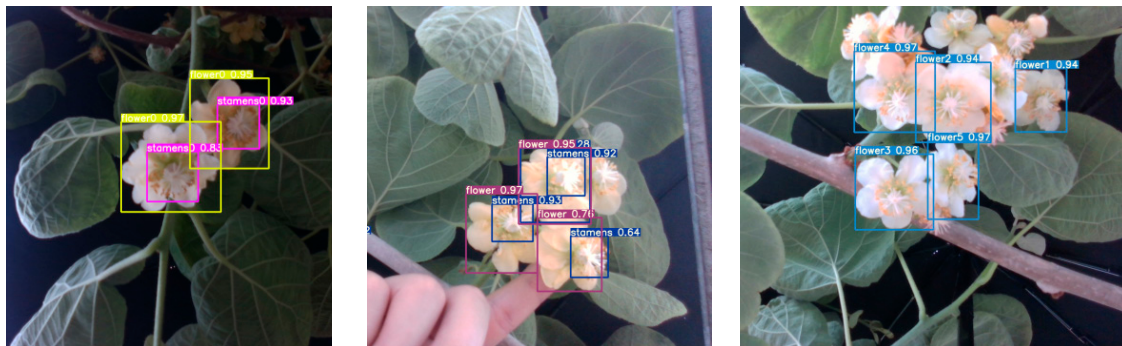
Figure 9. Test base.

Figure 10. Distribution of flower overlap.

A total of 880 kiwifruit flower images with good image quality were selected as the original data set, and the image data were manually labelled, including the flower and the core area of the same flower. The smallest external rectangle of the flower and the core area were used as the marking frame to ensure that the subject occupied a higher proportion of the marking frame. Flowers were labelled as the 'flower' class and stamens as the 'stamens' class; then, the images were converted to xml format to obtain more complete coordinate information. Data augmentation was performed with respect to the original data set, mainly in the form of rotation and the reduction and enhancement of brightness and Gaussian noise. The purpose of data augmentation is to enhance the useful information in an image, improve the visual properties of the image, purposefully emphasise the overall or local characteristics of the image, clarify the original image, or emphasise some interesting features. This process was conducted to distinguish different object features more effectively and improve the robustness of the model. Following data enhancement, a total of 3344 images were used as training data set used to train the recognition model. As shown in Figure 11.

Figure 11. Images after data enhancement.

Data sets D1, D2, D3, and D4 were constructed to compare the influence of data sets participating in the modelling of flower overlap recognition. For each overlap, 800 pieces were selected for model training, 200 pieces for model verification, and 200 pieces for model testing. The data set structure is shown in Table 1. Before the training procedure, the initial parameters of the YOLOv5 model were set (as shown in Table 2).

Table 1. Data set structure.

Data Set	Place	Condition of Overlap	Set of Training	Set of Verification	Set of Tests
D1		Single flower	800	200	200
D2	Orchard	Overlap of two	800	200	200
D3	greenhouses	Overlap of three	800	200	200
D4		Four or more overlapping	800	200	200

Table 2. Parameters for model training.

Parameters	Value
Input size/pixels	640×640
Initial learning rate	0.032
Momentum	0.843
Cyclical learning rate	0.12
Iteration	200

The size of the input image determines the amount of information calculated in the training process. An appropriate image size can achieve high training accuracy and reduce hardware loss in order to reduce computational work. According to preliminary studies, the size of the input image was selected as 640×640. Since VOC data set is used, the default parameter value obtained by optimization training applied to VOC data set is

selected. According to the data set size and the difficulty of target feature recognition, the number of iterations is set to 200.

The distribution of specific label information obtained after the training of the recognition model is shown in Figure 12. The flower image is of a moderate size due to the selection of an appropriate shooting distance. In the image of the recognition model, the flower, as a medium-sized target, enables the system to accurately identify its features. The dark-blue colour denotes the distribution area of labels. The size of labels is moderate, and the size of targets identified is relatively uniform. This data set is suitable for the close pollination of targets.
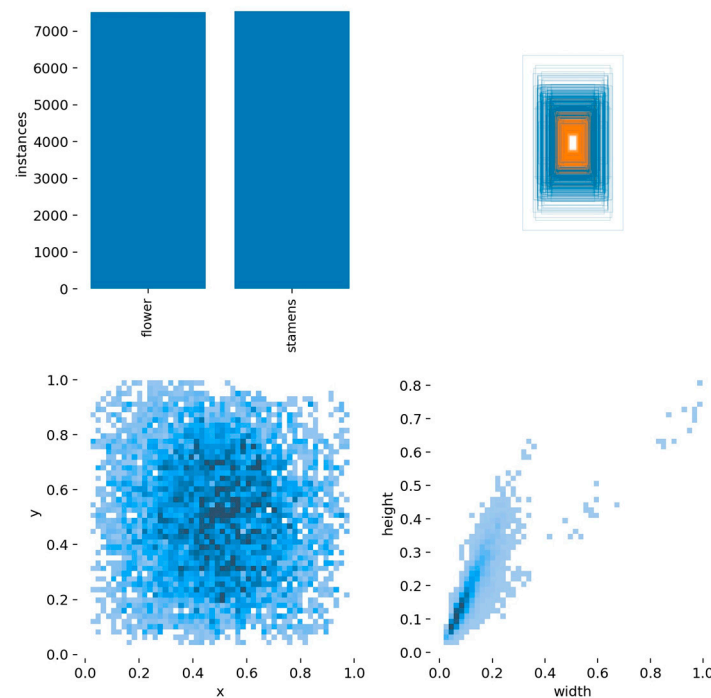
Figure 12. Label information distribution.

3. Results

3.1. Training Results

When collecting data on the kiwi flower distribution in a kiwifruit plantation, the number and overlap of flowers collected through the employed imaging system will be affected by changes in recognition distance. When the target pollination device is close to the kiwi flower, there are fewer flowers and flower overlap characteristics that can be recognised, but the overlap information of the kiwi flower can be identified more accurately. On the contrary, when the target pollination device is far from the kiwi flower, there will be more kiwi flowers in the visual system's recognition range, and there will also be more overlapping features of the flowers, which prevents the employed imaging device from effectively identifying the nature of the overlap of the kiwi flowers, thereby hampering the employed device's ability to identify the characteristics of the flowers. In the kiwi flower recognition training program relayed herein, the YOLOv5 with K-means++ and CBAM attention mechanism were used to identify kiwi flowers, and the optimized detect function was used to judge and calculate the flower's overlap and tilt angle.

A total of 800 training sets, 200 validation sets, and 200 test sets were used for each flower overlap. The test sets of flower overlap were divided into four categories: (1) single flower; (2) overlap of two flowers; (3) overlap of three flowers; and (4) overlap of four or more flowers. The training results are shown in Figure 13. As shown in Figure 13, the predicted value and recall rate tend to be stable at 100 rounds, with a maximum forecast confidence of 0.962, a maximum recall rate of 0.92, and a maximum mAP@0.5 value of 0.89.

From 100 rounds to 200 rounds, the above parameters are relatively stable and fluctuate only in a small range. Therefore, the model trained for 200 rounds was selected as the kiwi flower recognition model in the spray pollination process. The total number of kiwi flowers and the overlap number of the kiwi flowers can be obtained, and the accuracy, mAP value, maximum forecast confidence, and maximum recall rate can be calculated through this model.
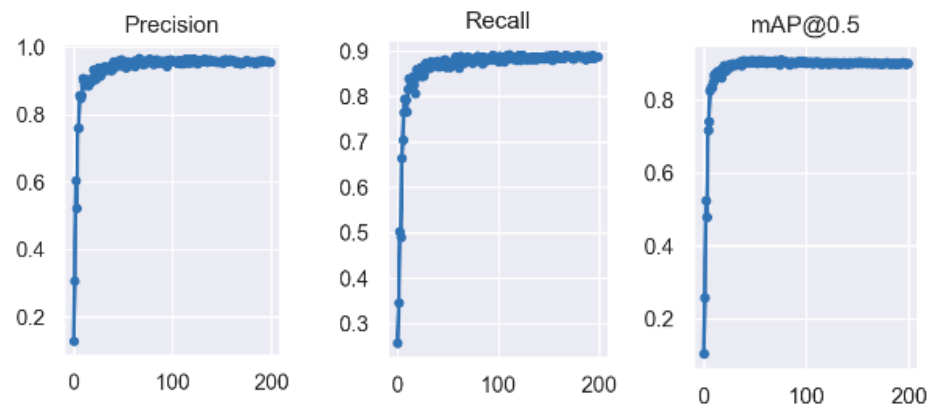

Figure 13. Prediction curve, Recall curve, and mAP value of the training results.

With regard to the visual system on the target device, its feature recognition accuracy can be maintained when the number of kiwifruit flowers is low. However, instances of false detection and failed detection will gradually increase when the number of flowers is increases and the degree of overlap becomes more complex. Therefore, the K-means++ algorithm and CBAM attention mechanism were incorporated into the YOLOv5 model, which can adjust the size of the anchor frame and select better clustering centres during training. The CBAM attention mechanism gives a higher weight to the overlapping features of kiwi flowers, so the visual system's ability to recognise the overlapping situation of flowers is enhanced, and a higher mAP value and maximum prediction confidence can be obtained compared with the results detected under the YOLOv5 model. These experimental results show that the optimised YOLOv5 model can more accurately identify the number of overlapping flowers and the tilt angle of flowers at higher quantities than those previously documented.

3.2. Accuracy Rate of Flower Angle Recognition

A certain error is observed between the actual tilt angle and the theoretical tilt angle of a kiwi flower due to the influence of the flower's own gravity. We calculated the regression function of the collected data and obtained a more accurate formula for calculating the tilt angle of the flower (see Formulae (4) and (5)). It was only necessary to identify the coordinates of the flower and the flower bud; then, the tilt angle of the flower could be calculated using Formula (5).

The flower tilt angle measured via image recognition was compared with the actual measured flower tilt angle, and the error rate (refer to Formula (6)) between the real tilt angle and the identified tilt angle was obtained through multiple groups of comparison experiments, as shown in Figure 14.

The tilt angles of 50 groups of flowers were measured using image recognition, and the data on the tilt angles of the two groups were compared. In accordance with the error rate chart of the tilt angle measurement of the flowers, the error between the tilt angle of the flowers measured by the calculation formula and the actual tilt angle was between 0.5% and 0.8%, which proved that the revised regression function tilt angle calculation formula was highly reliable. It can be used to measure the actual tilt angle of flowers, thereby providing accurate coordinate information that can be used to allow kiwi flowers to accurately pollinate the target and laying the foundations for effectively improving

the precision rate of pollination. The accuracy of flower tilt angles measured using this method is high, and the established model can accurately obtain the flower tilt angle with high reliability.
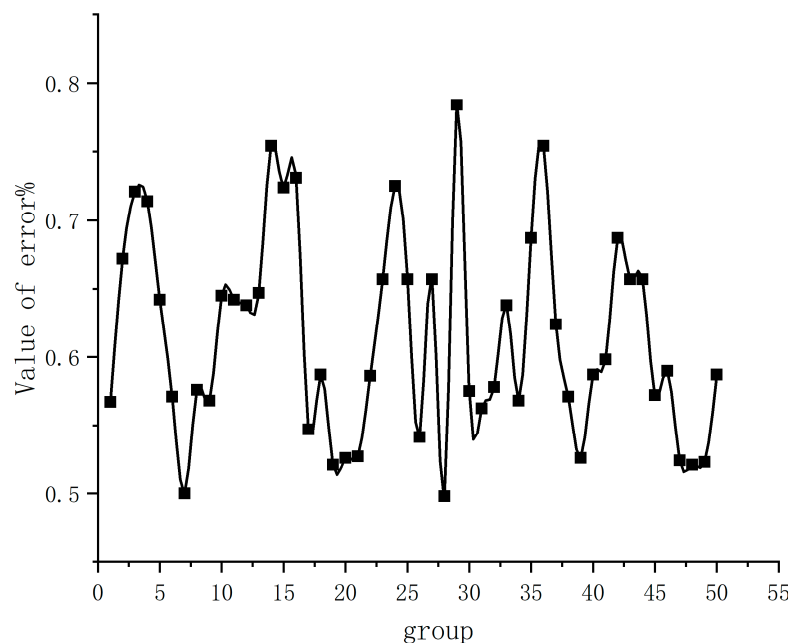

Figure 14. Error rate of flower tilt angle measurement.

3.3. Comparison Experiment Conducted to Identify Kiwi Flower Overlap

Kiwi flower overlap can be divided into four main conditions: single-flower overlap, two flowers overlapping, three flowers overlapping, and four and more flowers overlapping. The overlap of flowers is relatively complicated due to the complexity and variability of the shooting environments and angles. This condition leads to the failure of the employed visual system of the target device to accurately identify the features of the flowers and obtain the coordinates of the flowers and flower buds, which increases the probability of failed and false detection, reduces the accuracy of identification, and affects the success rate of pollination of flowers.

In this paper, YOLOv5 was used to identify the overlap of macaque peach blossoms. To compare the performance of different algorithms, such as Faster-RCNN, SSD, VGG, etc., in identifying flower overlap, the specific results of different target detection algorithm models are listed in Table 3.

Table 3. Test results of the five models regarding flower overlap.

Model	Predicted Value (Flowers)	Rate of Recall (Flowers)	Predicted Value (Stamens)	Rate of Recall (Stamens)	mAP@0.5
YOLOv5s	96.7%	89.1%	91.1%	78.3%	90.1%
Faster-RCNN-ResNet50	57.4%	98.9%	58.4%	97.9%	92.6%
Faster-RCNN-VGG	68.5%	98.9%	67.9%	98.0%	95.6%
SSD-VGG	76.6%	87.4%	82.8%	65.6%	82.3%
SSD-MobileNetv2	86.7%	70.2%	89.6%	55.0%	81.1%

Table 3 shows that the predicted accuracy of flowers and stamens obtained by YOLOv5s reached 96.7% and 91.1%, respectively, while maintaining high recall values. The predicted values of Fster-RCNN-ResNet50 and Faster-RCNN-VGG are both less than 70%, which is because the processes performed by the two algorithms are complex and their real-time performance is poor, which are both factors that are not conducive to the recognition of a large number of overlaps between flowers and stamens. Compared with YOLOv5, the

SSD-MobileNetv2 and SSD-VGG networks had deficiencies in terms of their small target identification features and extraction, and the predicted value decreased to less than 90%. In summary, in terms of overlapping target detection and small target detection, the overall ability of YOLOv5 is the best among the analysed models, and its predicted values, recall rate, and map@0.5 remain stable, indicating that the model can accurately identify kiwi flower overlap and obtain flower tilt angles.

The YOLOv5s model was used to conduct batch reasoning with the detection function for the recognition of flower overlap. The reasoning results are shown in Figure 15. As shown in Figure 15, the model we selected for kiwi flower overlap obtained good recognition results and can remarkably distinguish flowers from stamens. The model has a fast response speed, allowing it to identify the overlap of kiwi flowers and ensure a certain recognition speed. The YOLOv5s model can accurately calculate the centroid in accordance with the overlap situation and locate the pollination centre point after the modification of the detection function. In accordance with the calculation formula of the tilt angle, the tilt angle of the centre point of the edge of the overlapping polygon with respect to the target can be found. For multiple targets in the same picture, an overlap situation can be identified and judged simultaneously, which improves the efficiency of the subsequent targeting task; that is, it guarantees the accuracy of targeting.
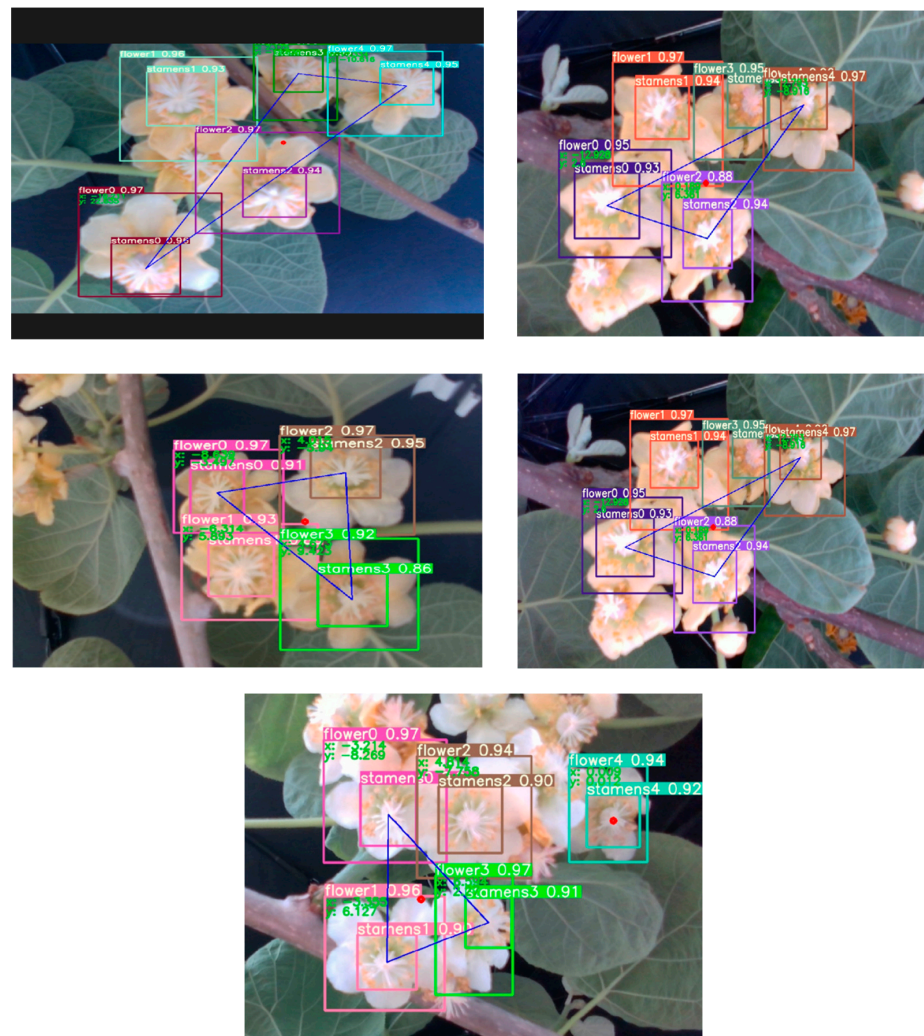

Figure 15. Inference result diagram of the detection function.

3.4. Comparative Test of Four YOLOv5 Models

An identification test was conducted for the four YOLOv5 series models to verify the rationality of the model selected in this study. The comprehensive performance of each model was determined to prove whether it was an optimal model. The test data of each model are shown in Table 4.

Table 4. Test results of the four models.

Model	F1	Average Time Per Frame (Milliseconds)	Memory (MB)
YOLOv5s	90.12%	8.64	20
YOLOv5m	93.97%	40.5	134
YOLOv5l	85.74%	425	278
YOLOv5x	91.24%	13.87	60.8

As shown in Table 4, all the models have high F1 rating values that surpassed 85%. The YOLOv5s model has the fastest average time per frame, reaching 8.64 milliseconds, which is higher than the other three. The YOLOv5s model only has 20 MB of memory, which is far smaller than the 100 MB or more of other models. The above data prove that YOLOv5s has high precision and a fast response speed, which is suitable for real-time detection used in small-scale agricultural operations.

Therefore, the proposed target identification technology of single-target and double-frame flower identification based on YOLOv5 can adequately realise the precise targeting of the pollination of kiwi flowers.

4. Discussion

In this study, a recognition model based on YOLOv5 was proposed for kiwi flower pollination. The experimental results show that the model has a high degree of recognition in terms of the overlap and tilt angle of kiwi flowers, can ensure extremely high accuracy, and can provide accurate coordinate information regarding kiwi flowers and flower buds, thus facilitating the development of an accurate pollination strategy.

In model's training process, the K-means++ clustering algorithm was adopted. Compared with the K-means algorithm, the optimised clustering algorithm can obtain a better initial clustering centre, select a more appropriate anchor frame size, and improve the calculation speed of the algorithm. The CBMA mechanism was incorporated into the model, which improved the model's extraction accuracy of kiwi flower features, effectively reduced the rate of missed detection and error detection, thus saving computational workload such that it was easier for the visual system to extract the features in the image. The optimization of the detection function improved the recognition of flower overlaps and the accuracy of flower tilt angle calculation and accurately determined the flower coordinates, pollination point coordinates, and pollination angles.

Different algorithms were used to identify the features of overlapping flowers. The effectiveness of YOLOv5 and other algorithms, such as Faster-RCNN, SSD, VGG, etc, was compared. Consequently, it was determined that the YOLOv5s model performs the best as the predicted values of flowers and stamens can reach 96.7% and 91.1%, respectively, while maintaining high recall values, which can meet the identification requirements regarding kiwi flower overlap and tilt angles. Among the four YOLOv5 series models, compared with YOLOv5m, YOLOv5l, and YOLOv5x, YOLOv5s has higher F1 values, the shortest average recognition time, and the smallest memory size, demonstrating its advantages of a fast response speed and high precision.

The algorithmic model of this study is suitable for kiwifruit plantations with flat ground but requires suitable lighting conditions. For kiwifruit plantations with rugged ground, it is very likely that smaller targets may appear, so a small target detection layer needs to be incorporated into the algorithm model to improve its recognition accuracy. Moreover, considering the night pollination to improve overall pollination efficiency, it is

necessary to enhance the illumination and corresponding filtering treatment of the images selected at night to improve the accuracy of night target recognition.

References

1. Guan, L. Japan's kiwi artificial pollination technology. *J. Deciduous Fruit Trees* **2002**, *5*, 60.
2. Jiang, Z.J. Optimization of Double-Flow Spray Parameters and Development of Kiwifruit Pollination Device. Master's Thesis, Northwest Agriculture and Forestry University of Science and Technology, Xi'an, China, 2020.
3. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
4. Liu, T.; Teng, G.; Yuan, Y.; Liu, B.; Liu, Z. Recognition method of winter jujube fruit in natural scene based on improved YOLO-v3. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 17–25.
5. Zhang, H.; Fu, Z.; Han, W.; Yang, G.; Niu, D.; Zhou, X. Maize seedling number acquisition method based on improved YOLO. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 221–229.
6. Li, K.J. Research and Application of Weed Detection Algorithm in Cotton Field based on YOLOv3. Master's Thesis, Xinjiang University, Ürümqi, China, 2021.
7. Yue, Y.; Li, X.; Zhao, H.; Wang, H. Crop disease image recognition based on improved VGG networks. *J. Agric. Mech. Res.* **2022**, *44*, 18–24.
8. Li, H.; Li, C.; Li, G.; Chen, L. A real-time table grape detection method based on improved YOLOv4-tiny network in complex background. *Biosyst. Eng.* **2021**, *212*, 347–359. [CrossRef]
9. Wang, H.; Xu, Y.; He, Y.; Cai, Y.; Chen, L.; Li, Y.; Angel Sotelo, M.; Li, Z. YOLOv5-Fog: A Multiobjective Visual Detection Algorithm for Fog Driving Scenes Based on Improved YOLOv5. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–12. [CrossRef]
10. Lv, J.; Xu, H.; Han, Y.; Lu, W.; Xu, L.; Rong, H.; Yang, B.; Li, Z.; Ma, Z. A visual identification method for the apple growth forms in the orchard. *Comput. Electron. Agric.* **2022**, *197*, 106954. [CrossRef]
11. Chaschatzis, C.; Karaiskou, C.; Mouratidis, E.G.; Karagiannis, E.; Sarigiannidis, P.G. Detection and Characterization of Stressed Sweet Cherry Tissues Using Machine Learning. *Drones* **2021**, *6*, 3. [CrossRef]
12. Ficzere, M.; Mészáros, L.A.; Kállai-Szabó, N.; Kovács, A.; Antal, I.; Nagy, Z.K.; Galata, D.L. Real-time coating thickness measurement and defect recognition of film coated tablets with machine vision and deep learning. *Int. J. Pharm.* **2022**, *623*, 121957. [CrossRef]
13. Jin, Y.; Gao, H.; Fan, X.; Khan, H.; Chen, Y. Defect Identification of Adhesive Structure Based on DCGAN and YOLOv5. *IEEE Access* **2022**, *10*, 79913–79924. [CrossRef]
14. Li, H.; Yang, G. Dietary Nutritional Information Autonomous Perception Method Based on Machine Vision in Smart Homes. *Entropy* **2022**, *24*, 868. [CrossRef]
15. Xue, J.; Zheng, Y.; Dong-Ye, C.; Wang, P.; Yasir, M. Improved YOLOv5 network method for remote sensing image-based ground objects recognition. *Soft Comput.* **2022**, *26*, 10879–10889. [CrossRef]
16. Yao, J.; Qi, J.; Zhang, J.; Shao, H.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. *Electronics* **2021**, *10*, 1711. [CrossRef]
17. Zhao, W.; Liu, S.; Li, X.; Han, X.; Yang, H. Fast and accurate wheat grain quality detection based on improved YOLOv5. *Comput. Electron. Agric.* **2022**, *202*, 107426. [CrossRef]
18. Zhang, H.; Tian, M.; Shao, G.; Cheng, J.; Liu, J. Target Detection of Forward-Looking Sonar Image Based on Improved YOLOv5. *IEEE Access* **2022**, *10*, 18023–18034. [CrossRef]
19. Zhang, Z.; Qiao, Y.; Guo, Y.; He, D. Deep Learning Based Automatic Grape Downy Mildew Detection. *Front. Plant Sci.* **2022**, *13*, 872107. [CrossRef]

20. Dai, G.; Hu, L.; Fan, J.; Yan, S.; Li, R. A Deep Learning-Based Object Detection Scheme by Improving YOLOv5 for Sprouted Potatoes Datasets. *IEEE Access* **2022**, *10*, 85416–85428. [CrossRef]

21. Gao, G.; Wang, S.; Shuai, C.; Zhang, Z.; Zhang, S.; Feng, Y. Recognition and Detection of Greenhouse Tomatoes in Complex Environment. *Traitement Signal.* **2022**, *39*, 291–298. [CrossRef]

22. Ye, Z.; Guo, Q.; Wei, J.; Zhang, J.; Zhang, H.; Bian, L.; Guo, S.; Zheng, X.; Cao, S. Recognition of terminal buds of densely-planted Chinese fir seedlings using improved YOLOv5 by integrating attention mechanism. *Front. Plant Sci.* **2022**, *13*, 991929. [CrossRef]

23. Li, D.H.; Zhao, H.; Yu, X. Overlapping green apple recognition based on improved spectral clustering. *Spectrosc. Spect. Anal.* **2019**, *39*, 2974–2981.

24. Uijlings, J.R.; Van De Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]

25. Miao, R.; Shan, Z.; Zhou, Q.; Wu, Y.; Ge, L.; Zhang, J.; Hu, H. Real-time defect identification of narrow overlap welds and application based on convolutional neural networks. *J. Manuf. Syst.* **2022**, *62*, 800–810. [CrossRef]