applied sciences

MDPI

Article

Flowering Index Intelligent Detection of Spray Rose Cut Flowers Using an Improved YOLOv5s Model

Junyan Li [1,*] and Ming Li [2]

[1] College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China
[2] School of Electrical Engineering, Anhui Polytechnic University, Wuhu 214000, China; swfu_lm@swfu.edu.cn
* Correspondence: 19422107@stu.njau.edu.cn; Tel.: +86-15087010923

Abstract: Addressing the current reliance on manual sorting and grading of spray rose cut flowers, this paper proposed an improved YOLOv5s model for intelligent recognition and grading detection of rose color series and flowering index of spray rose cut flowers. By incorporating small-scale anchor boxes and small object feature output, the model enhanced the annotation accuracy and the detection precision for occluded rose flowers. Additionally, a convolutional block attention module attention mechanism was integrated into the original network structure to improve the model's feature extraction capability. The WIoU loss function was employed in place of the original CIoU loss function to increase the precision of the model's post-detection processing. Test results indicated that for two types of spray rose cut flowers, Orange Bubbles and Yellow Bubbles, the improved YOLOv5s model achieved an accuracy and recall improvement of 10.2% and 20.0%, respectively. For randomly collected images of spray rose bouquets, the model maintained a detection accuracy of 95% at a confidence threshold of 0.8.

Keywords: spray rose; YOLOv5; flowering index; cut flower; detection

check for updates

Citation: Li, J.; Li, M. Flowering Index Intelligent Detection of Spray Rose Cut Flowers Using an Improved YOLOv5s Model. *Appl. Sci.* **2024**, *14*, 9879. https://doi.org/10.3390/app14219879

Academic Editor: João M. F. Rodrigues

Received: 12 September 2024
Revised: 25 October 2024
Accepted: 26 October 2024
Published: 29 October 2024

1. Introduction

With the ongoing enhancement of living standards, the demand for cut roses has been consistently on the rise. In recent years, spray roses (*Rosa rugosa Thunb.*) have gained significant popularity in the market due to their distinctive characteristic of "one stem, multiple flowers". According to statistics, in 2022, the Kunming International Flora Auction Trading Center (KIFA) auctioned a total of 251 varieties of spray roses. This represents an increase of 21 varieties compared to 2021, resulting in an overall supply growth of 20.2%. Notably, orange and yellow spray roses comprised 53% of the total supply. Grading detection is a critical process in the quality assessment of cut roses, playing an essential role in ensuring both market competitiveness and planting efficiency for rose cultivation. The grading of cut flowers must take into account various parameters, including the characteristics of both the flower and stem, with the degree of flower opening serving as a crucial indicator for assessment. While existing standards have established methods for evaluating the openness of cut flowers, the absence of clear and quantifiable measurement criteria means that the grading process for spray rose openness continues to rely significantly on manual inspection.

In recent years, the rapid advancement of artificial intelligence technology has positioned deep learning-based object detection as a prominent area of research. Convolutional neural network (CNN) models for object detection can be broadly categorized into two types: two-stage detection models that rely on candidate regions, such as RCNN [1], Faster R-CNN [2], and Mask R-CNN [3], and one-stage detection models based on regression techniques, including YOLO [4] and SSD [5]. One-stage detection models bypass the candidate region generation step by directly extracting features, making them increasingly prevalent in modern intelligent agriculture. Various versions of You Only Look Once (YOLO) have been applied to tasks such as plant disease identification, pest detection, and fruit ripeness assessment. For instance, Mathew et al. utilized the YOLOv5 model to detect bacterial spot disease in the leaves of bell pepper plants [6]. Soeb et al. employed the YOLOv7 model for identifying leaf diseases in tea leaves, achieving a detection accuracy of 97.3% [7]. Xue et al. proposed an enhanced YOLOv5 model aimed at detecting diseases and insect pests affecting tea leaves. This improvement integrated a convolutional block attention module along with self-attention and convolution techniques into the YOLOv5s framework to enhance feature image identification capabilities [8]. Tian et al. introduced a Multi-scale Dense YOLO model designed for detecting three typical small target lepidopteran pests on sticky insect boards, with experimental results indicating that the mAP@0.5 value reached 86.2% [9]. Addressing performance limitations in existing models for small target detection, Wu et al. incorporated the AFFD Head detection module into the YOLOv8 architecture to identify small citrus pests; this improved model achieved an average accuracy of 90.3% [10]. Qin et al. also presented an enhanced YOLOv8 model specifically for detecting soybean insect pests, attaining a precision rate of 80.8% [11].

In terms of fruit detection, Zhao et al. developed an improved YOLOv5 model capable of identifying five growth stages of pomegranate: bud, flower, early-fruit, mid-growth, and ripe phases; this modified version demonstrated a detection accuracy of 92.2%. Although this figure was lower than that obtained by the original YOLOv5s model, it exhibited a speed increase of 17.3% compared to its predecessor [12]. Lin et al., on their part, proposed an attention-based YOLO network designed to detect both unripe and ripe citrus fruits; their findings indicated that precision could reach as high as 90.6% [13]. By substituting the backbone component of YOLOv5 with a lightweight Ghost Net module, Sun et al.'s lightweight YOLO model effectively detected passion fruit within complex environments—achieving an average accuracy rate of 96%, which is notably 1% higher than that recorded by the original version [14]. Shi et al. introduced a lightweight YOLOv8 model for the detection and quantification of peach fruits, achieving a detection precision of 97.9% [15]. Wu et al. proposed an MTS-YOLO model by reconstructing the neck component of YOLOv8 to identify the maturity and picking positions of tomatoes. Experimental results demonstrated that this enhanced YOLO model attained an F1-score of 88.7% and a mAP@0.5 of 92.0% [16].

The application of machine learning to the classification of cut flowers can be traced back to 1994 when Steinmetz et al. developed a bud maturity grading classifier utilizing Bayesian decision methods based on manually labeled classification results [17]. Hiary et al. introduced a classification approach that employed a fully convolutional neural network (FCN) alongside a pre-trained Visual Geometry Group (VGG) model, achieving an impressive recognition accuracy of 97% on a flower dataset [18]. Mete et al. integrated features extracted by convolutional networks with a support vector machine classifier, leveraging deep convolutional networks and machine learning algorithms for the recognition and classification of flower images [19]. Anjani et al. utilized CNNs for the classification of rose flowers, attaining a recognition accuracy of 96.33% for RGB images of flowers [20]. Sun et al. employed four CNNs—VGG16, ResNet18, MobileNetV2, and InceptionV3—to assess the maturation status of individual fresh-cut roses. The findings indicated that the grading accuracy of the enhanced InceptionV3-based model reached as high as 98% [21].

Recently, models from the YOLO family have been widely utilized for flower identification and grading. It is well established that both the quality and quantity of flowers and leaves can serve as indicators of crop growth. Notably, counting the number of flowers in crops significantly aids in improving growers' management practices. Consequently, a series of YOLO models have been implemented to detect flowering crops. For example, Wang et al. utilized an improved YOLOv8 model to identify and count flower buds and blossoms on rapeseed (*Brassica napus*) [22]. Oishi et al. applied YOLOv3 and Faster R-CNN models to classify abnormal versus healthy potato leaves based on images captured by a portable video camera; their results demonstrated that compared to YOLOv3, the Faster R-CNN model exhibited superior robustness under varying lighting conditions [23].

Bai et al. introduced an enhanced YOLOv7 model designed for accurate identification of flowers and fruits in strawberry seedlings within greenhouse environments; this was achieved by incorporating a high-resolution prediction head into the original architecture to improve detection capabilities for small, overlapping, or occluded targets [24]. Qi et al. proposed a highly fused yet lightweight detection framework known as Fusion-YOLO aimed at identifying early flowering stages in tea chrysanthemums; this improved version adapts effectively to variations in illumination along with occlusion and overlap scenarios [25]. Zhou et al. developed S-YOLO—a model tailored for monitoring apple tree flowering periods within modern standardized orchards—by substituting its backbone network with Swin Transformer-tiny within an upgraded YOLO framework [26]. Wu et al. employed a channel pruning-based YOLOv4 model to detect apple flowers for automatic thinning in natural environments. A channel pruning algorithm was integrated into the foundational YOLOv4 model to streamline the detection process [27]. Wang et al. utilized an enhanced YOLOv4 model to identify potted flowers and their spatial locations [28]. Lyu et al. introduced a detection method for male and female litchi flowers based on YOLO models, with the objective of achieving a balanced control of their ratio. They used both the YOLOv4 and YOLOv5 models as teacher models while leveraging a relatively simpler YOLOv4-Tiny as the student model [29]. Duan et al. proposed a real-time, high-precision end-to-end sorting system based on the YOLO network for fresh-cut flowers, capable of performing flower localization, classification, and grading tasks [30]. Zhang et al. proposed an enhanced YOLOv8s model for the grading and detection of cut roses, with experimental comparisons indicating that the precision of this improved model reached 97.4% [31]. To the best of our knowledge, there are currently no reported applications of machine learning in grading detection for spray rose cut flowers.

Unlike the grading of roses, spray roses can exhibit flowers at varying degrees of openness on the same stem. Consequently, when assessing spray roses, it is essential to evaluate and categorize each flower according to its flowering index. Taking into account the unique characteristics of spray roses while balancing network complexity with deployment capabilities, this study has selected YOLOv5s as the foundational model for detecting and grading the flowering index of spray rose cut flowers. In light of the distribution characteristics inherent in spray roses, several enhancements have been proposed, including improved small object detection, integration of attention mechanisms, and modifications to anchor box settings and loss functions. The trained network model was subsequently deployed on a mobile device to present detection results for rose bouquets and calculate the proportions of various flowering indices among cut spray roses. This approach provides a solid foundation for effectively sorting spray rose cut flowers based on their flowering index.

2. Materials and Methods

2.1. Image Acquistion

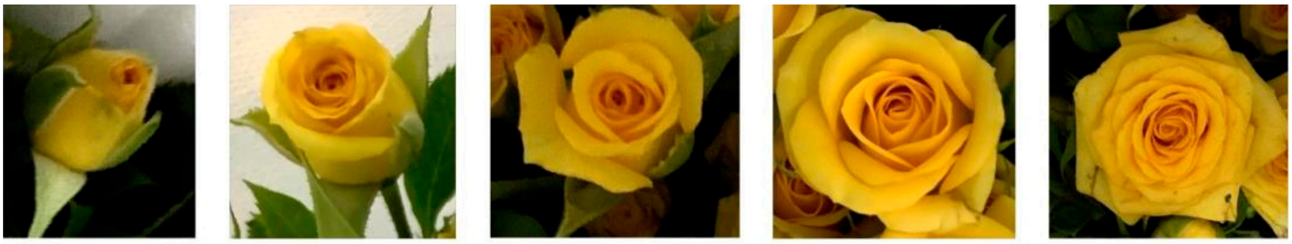
This study focused on two varieties of spray rose cut flowers—Orange Bubbles and Yellow Bubbles—within the orange and yellow color series. The images were collected at the Jinhai Flower Base in Tonghai County, Yunnan Province, which is the primary rose cultivation area in Yunnan. To enhance the diversity of the image sample set, spray rose cut flowers were photographed using both a Canon EOS 600D (Canon Inc., Tokyo, Japan) digital camera and a HUAWEI Mate 40 smartphone (Huawei Technologies Co., Ltd., Shenzhen, China). A total of 2145 images of roses with varying degrees of flower openness were selected to construct the dataset for network training and testing.

The degree of flower openness serves as a crucial grading indicator for cut flowers and provides the most immediate visual impression for consumers. The national standard categorizes rose flower openness into five distinct levels, each accompanied by a qualitative description (Table 1).

Table 1. Standards for rose flowering index (GB/T 41201-2021) [32].

Flowering Index	Description
1	Sepals remain upright, petals tightly closed, not extending beyond the sepals
2	Sepals open to 30–45 degrees, outer petals begin to loosen
3	Sepals open to 45–90 degrees, outer petals loosen and spread
4	Sepals slightly droop, outer petals curl outward, multiple layers of petals spread
5	Petals are fully loose, multiple layers of petals curl

Due to the lack of quantitative evaluation indicators, the grading of flower openness in spray roses currently relies on manual identification. Figure 1 shows the manually labeled results corresponding to the different flowering index, denoted as Fi-*j*, where *j* = 1, 2, 3, 4, 5. Furthermore, since the two types of spray rose cut flowers studied were primarily sold outside the province and required a longer transportation time, their flowering indices predominantly fall within levels 1 to 3. Consequently, among the images collected for both color series, 59.4% of samples exhibited flowering indices at levels 1 and 2; conversely, level 5 accounted for only 8.1%. The specific quantity and distribution of the dataset are shown in Table 2.



Fi-1 Fi-2 Fi-3 Fi-4 Fi-5

Figure 1. Manual labeling of the different flowering index.

Table 2. Quantity and distribution of the spray rose dataset.

Spray Rose	Fi-1	Fi-2	Fi-3	Fi-4	Fi-5
Yellow Bubbles	416	314	238	152	98
Orange Bubbles	322	230	183	124	76
Total	734	540	421	276	174
Proportion (%)	34.2	25.2	19.6	12.9	8.1

All rose flower samples in the dataset were meticulously labeled for flowering index according to the standards outlined in Table 1 and illustrated in Figure 1, utilizing Labeling software (labelImg 1.8.4). The labels are designated as O-Fi-j and Y-Fi-j, where j = 1, 2, 3, 4, 5. The capital letters O and Y represent the orange and yellow color series of spray roses, respectively. Data augmentation techniques, including Cutout, Add Noise, Change Light, Translate, Rotate, and Mirror, were then applied to increase the dataset size by eight times (see Figure 2). Finally, the image dataset was randomly divided into three groups: training set, validation set, and test set, in a 7:2:1 ratio.

2.2. Improved YOLOv5s Network Model

The YOLO series of models is a fast and efficient object detection algorithm that treats the object detection task as a regression problem. Unlike traditional object detection algorithms, YOLO uses a single neural network to simultaneously predict the bounding boxes and classes of multiple objects within an image. Compared to its predecessors, YOLOv5 maintains high accuracy while offering smaller weight files and faster training and inference speeds. In addition, Duan et al. conducted a comparative analysis of the detection performance among four YOLO models: YOLOv5, YOLOv6, YOLOv7, and YOLOv8, specifically in grading single fresh-cut roses. The Average Precision (AP) for these models was recorded at 99.98%, 96.87%, 100%, and 99.99%, respectively; however, their processing speeds (measured in Frames Per Second or FPS) were noted as 76.49, 56.47, 23.55, and 60.85 correspondingly. In terms of complexity, the number of parameters for YOLOv5, YOLOv6, YOLOv7, and YOLOv8 was 7.02M, 18.50M, 36.50M, and 11.13M, respectively. It was evident that the YOLOv5 model was the simplest among them. Although the precision of YOLOv8 exceeded that of the YOLOv5 model by 0.01%, it came at a cost—its parameter scale was 58.55% higher. This outcome indicated that while the precision of the YOLOv5 model was comparable to that of the other three models, its complexity and speed were significantly superior [30]. The results obtained indicated that the inference times for YOLOv5s and YOLOv8s were 5.9 ms and 15.6 ms, respectively, when applied to the same scene. This suggested that the detection speed of YOLOv5s was more than three times faster than that of the YOLOv8s model [31]. The YOLOv5 model offered advantages in terms of speed and simplicity, which align well with the requirements of this research. Therefore, we selected the construction of the YOLOv5 model for grading the flowering index of spray rose cut flowers. YOLOv5 is primarily composed of four parts: Input, Backbone, Neck, and Head. The Input layer uses the mosaic method for data augmentation, performs adaptive anchor box calculations, and applies mirror normalization to the input images. The Backbone layer extracts features at different scales from the images for use in subsequent network layers. The Neck layer fuses the multi-scale features extracted by the Backbone and passes the fused feature maps to the Head layer for prediction. The Head layer employs the CIoU loss function for bounding box regression and uses Non-Maximum Suppression (NMS) to eliminate redundant detection boxes.

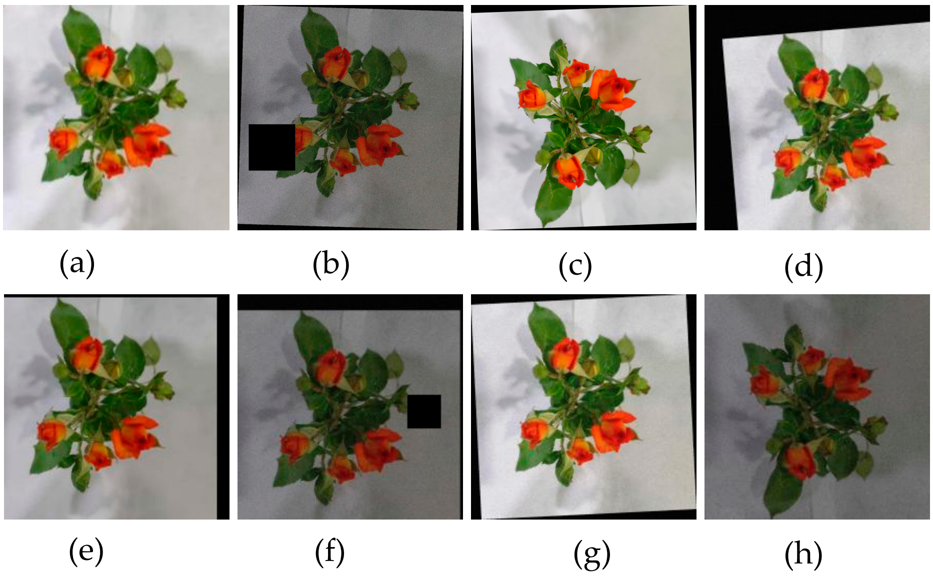


Figure 2. Images after data augmentation. (**a**) Original; (**b**) Cutout and Add Noise; (**c**) Change Light; (**d**) Translate and Rotate; (**e**) Translate; (**f**) Translate and Cutout; (**g**) Add Noise; (**h**) Mirror and Add Noise.

Considering the trade-off between detection accuracy, computational cost, and model complexity, YOLOv5s was selected as the baseline framework for optimization and improvement, with the network structure shown in Figure 3. The specific improvement strategies included the following four aspects: (1) In addition to the three-scale feature maps output by the detection head in YOLOv5s, an additional small-scale detection output was added to improve the detection accuracy of the flowering index of small, occluded flowers within the bouquet; (2) a spatial and channel attention mechanism was incorporated to enhance the feature extraction capability for the flowering index; (3) the WIoU loss function replaced the CIoU loss function utilized by YOLOv5s to better accommodate varying scales and aspect ratios of rose targets while improving bounding box regression accuracy across different levels; (4) for optimizing detection results, DIoU-NMS was used instead of the standard NMS in YOLOv5s to enhance the detection accuracy of densely distributed roses within the bouquet.

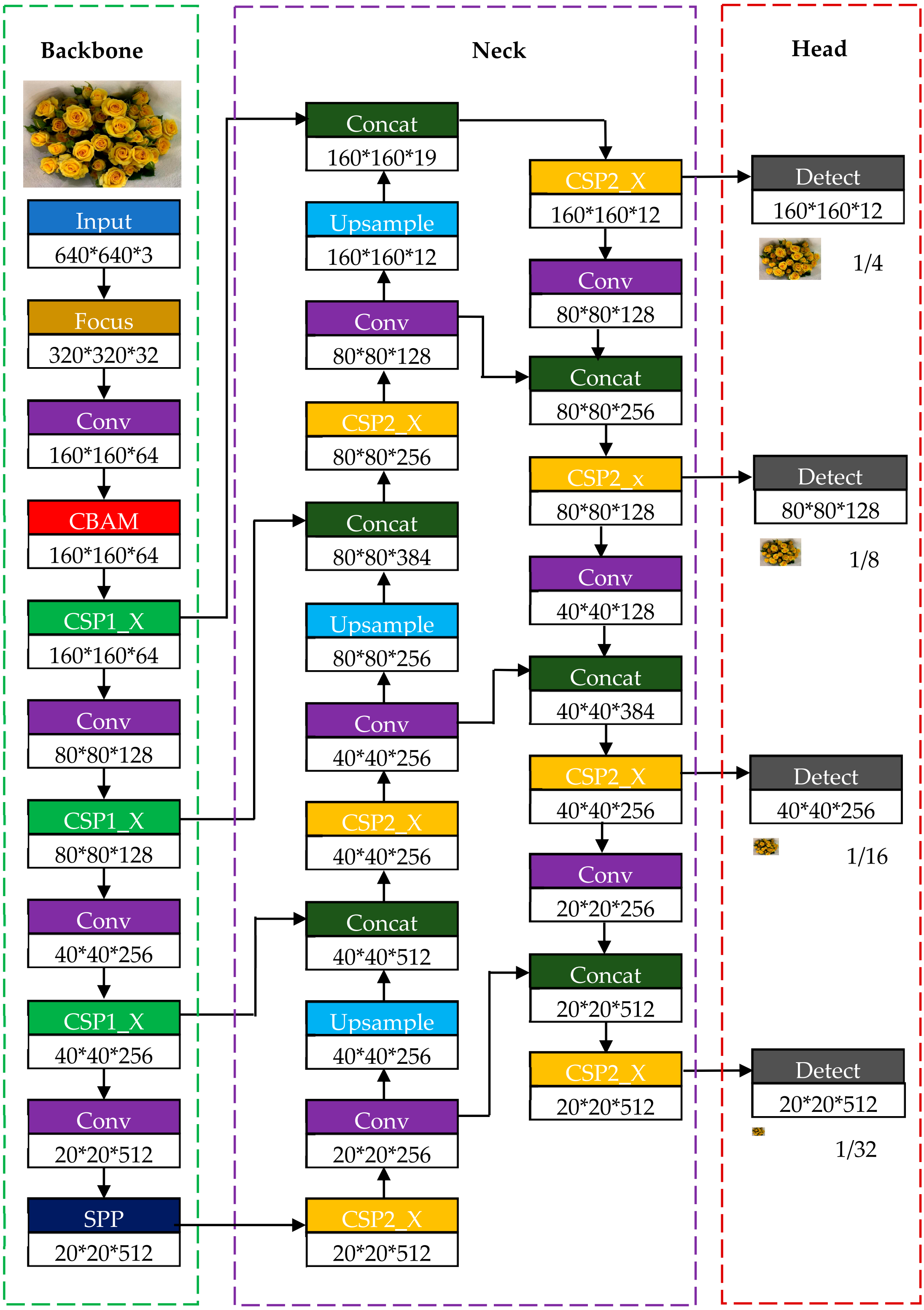


Figure 3. Improved YOLOv5s model network structure. Note: CSP1_X refers to the first CSP module; CSP2_X refers to the second CSP module.

2.3. Small-Scale Detection and Anchor Box Configuration

The original YOLOv5 network outputs three feature maps of different sizes at the Head layer. For an input image of 640 × 640, YOLOv5 produces feature maps at scales of 80 × 80, 40 × 40, and 20 × 20 (see Figure 3) to achieve multi-scale object detection. In the images of rose bouquets captured in this study, small-scale flowers were often formed due to occlusion. To address this, an additional feature map output at a scale of 160 × 160 was added to the Head layer in the improved network. Specifically, the 80 × 80 feature map was upsampled to 160 × 160, then concatenated with the feature map obtained from the second step of the Backbone layer, forming an additional output at the Head layer.

The initial anchor boxes in YOLOv5 are set based on the COCO dataset, which differs significantly in target size compared to the rose dataset used in this study. Therefore, the K-Means clustering algorithm was employed to adaptively calculate the anchor box sizes based on the annotated target box dimensions. Additionally, to accommodate the improved small-scale feature map output, an extra set of anchor boxes was added. The distribution of the modified anchor box sizes is shown in Table 3.

Table 3. Quantity and distribution of the spray rose dataset.

Scale	20 × 20	40 × 40	80 × 80	160 × 160
Receptive field	Large	Middle	Small	Very small
Sizes of anchors	116 × 90 156 × 198 373 × 326	30 × 61 62 × 45 59 × 119	10 × 13 16 × 30 33 × 23	5 × 6 8 × 14 15 × 11

2.4. Convolutional Block Attention Module (CBAM) Attention Mechanism

In images of rose bouquets, roses often exhibited a concentrated distribution in certain areas. To enhance the feature learning capability, the CBAM attention mechanism was introduced [18]. CBAM consists of two independent sub-modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM) (see Figure 4). CAM helps the model determine which channel features should be prioritized, while SAM aids the model in identifying the most meaningful spatial locations to focus on. The process for generating the CAM weight coefficients is described by the following equation:

$$\begin{aligned}\mathbf{M_c}(\mathbf{F}) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(\mathbf{F})) + \mathrm{MLP}(\mathrm{MaxPool}(\mathbf{F}))) \\ &= \sigma\left(\mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{avg}}\right)\right) + \mathbf{W_1}(\mathbf{W_0}(\mathbf{F^c_{max}}))\right)\end{aligned} \tag{1}$$

where $\mathbf{F}$ is the input feature map, σ is the sigmoid activation function, *MLP* is a two-layer perceptron, *AvgPool* and *MaxPool* represent average pooling and max pooling, respectively, $\mathbf{F^c_{avg}}$ and $\mathbf{F^c_{max}}$ denote the feature maps obtained by average and max pooling, and $\mathbf{W_1}$ and $\mathbf{W_2}$ are the weights of the two-layer *MLP*.

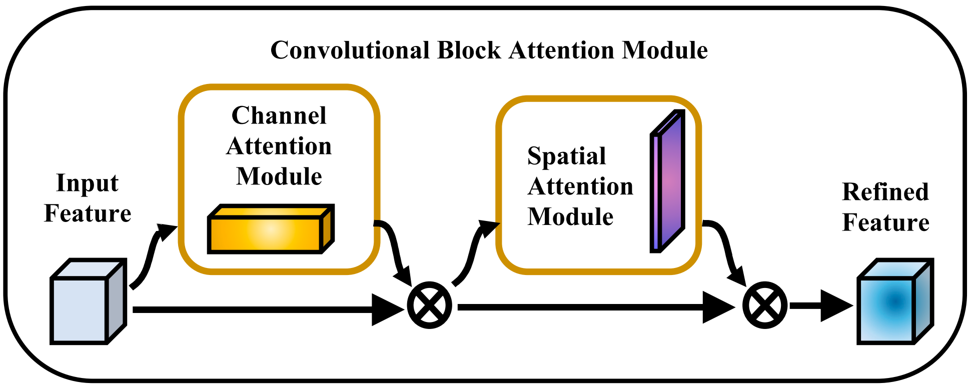


Figure 4. Structure of the CBAM Attention Mechanism.

Finally, element-wise multiplication of $\mathbf{M_c}(\mathbf{F})$ and the input feature map $\mathbf{F}$ yields the output feature map of the CAM.

The process for generating the SAM weight coefficients is as follows:

$$\begin{aligned}\mathbf{M_s}(\mathbf{F}) &= \sigma\left(f^{7\times7}([\mathrm{AvgPool}(\mathbf{F});\mathrm{MaxPool}(\mathbf{F})])\right)\\ &= \sigma\left(f^{7\times7}\left(\left[\mathbf{F^s_{avg}};\mathbf{F^s_{max}}\right]\right)\right)\end{aligned} \tag{2}$$

where $\mathbf{F}$ is the feature map output by the CAM module, and $f^{7\times7}$ denotes a 7×7 convolution operation. For the input feature map $\mathbf{F}$, a channel-based max and average pooling operation is first performed, producing two H × W × 1 feature maps. These two feature maps are then concatenated channel-wise and passed through a 7 × 7 convolution operation, resulting in an H × W × 1 feature map. The sigmoid function generates the SAM weight coefficients and $\mathbf{M}_S(\mathbf{F})$ similar to CAM, and element-wise multiplication with the input feature map yields the output feature map of SAM.

2.5. WIoU Loss Function

In object detection tasks, the similarity between the predicted bounding box and the ground truth bounding box is a crucial evaluation metric. The traditional Intersection over Union (IoU) loss function is widely used to measure the overlap between the predicted and ground truth bounding boxes. In Figure 5, Ws and Hs represent the width and height of the smallest enclosing box, while Wi and Hi represent the width and height of the overlap between the predicted and ground truth bounding boxes. The red line represents the distance between the center points the two boxes. The green area represents the predicted bounding box B = (x, y, w, h), and the yellow area represents the ground truth bounding box B_{gt} = (x_{gt}, y_{gt}, w_{gt}, h_{gt}). The IoU_Loss is defined as follows:

$$\mathcal{L}_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{S_u} \tag{3}$$

where S_u refers to the union of the predicted and ground truth boxes. However, when there is no overlap between the predicted and ground truth bounding boxes, the gradient of the L_{IoU} during backpropagation becomes zero. To address this, the original YOLOv5 model employs a CIoU_Loss function, which includes a penalty term:

$$\mathcal{L}_{CIoU} = \mathcal{L}_{IoU} + \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_s^2 + H_s^2} + \alpha v \tag{4}$$

$$\begin{aligned}\alpha &= \tfrac{v}{v+\mathcal{L}_{IoU}}\\ v &= \tfrac{4}{\pi^2}\left(\arctan\tfrac{w}{h} - \arctan\tfrac{w_{gt}}{h_{gt}}\right)\end{aligned} \tag{5}$$

where α is a weighting function, and ν is a similarity measure for the aspect ratio. Since CIoU_Loss only reflects the aspect ratio differences, this paper introduced the WIoU_Loss [19] as the bounding box loss function to better evaluate the actual relationship between the predicted and ground truth boxes. The key idea behind this improvement is to incorporate a distance attention mechanism into the IoU_Loss. The WIoU_Loss is calculated as follows:

$$\begin{aligned}\mathcal{L}_{WIoU} &= \mathcal{R}_{WIoU}\mathcal{L}_{IoU}\\ \mathcal{R}_{WIoU} &= \exp\left(\tfrac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_s^2+H_s^2)^*}\right)\end{aligned} \tag{6}$$

where * denotes the separation of W_s and H_s from the computation graph. The improved WIoU Loss function can focus on anchor boxes of normal quality and enhance the overall detection performance.

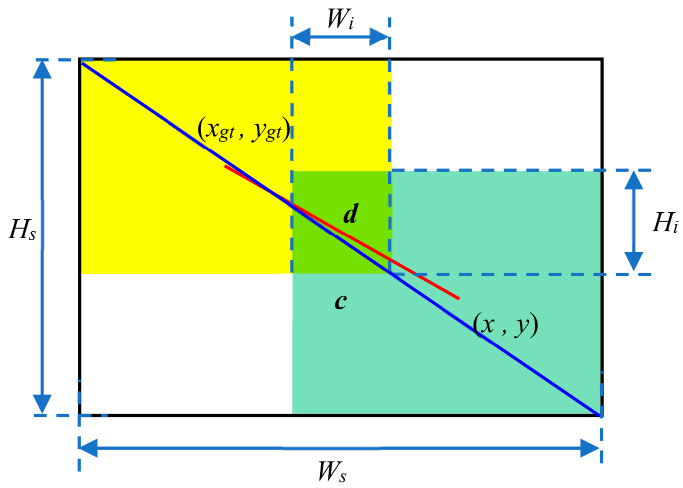


Figure 5. Illustration of predicted and ground truth bounding boxes.

2.6. DIoU-NMS Algorithm

NMS is a post-processing algorithm used in YOLOv5 for object detection, primarily to eliminate redundant detection boxes. Traditional NMS employs IoU as the evaluation metric, relying solely on the overlap area between two boxes to discard redundant detection boxes. However, in images of rose bouquets, two roses with the same degree of openness may be closely adjacent, causing NMS to erroneously classify one of the detection boxes as redundant and subsequently remove it. To address this, the improved DIoU-NMS algorithm is used as the post-processing method for object detection. This algorithm not only considers the overlap area of the detection boxes but also incorporates the distance between the center points of the two boxes as an evaluation metric. For pairs of boxes with a high IoU value, if there exists a significant distance between their center points, they are regarded as distinct objects and thus retained rather than deleted. The formula for the DIoU-NMS algorithm is as follows:

$$\begin{gathered} \widetilde{s_i} = \begin{cases} s_i, IoU - \mathcal{R}_{DIoU}(\mathcal{M}, B_i) < \varepsilon \\ 0, IoU - \mathcal{R}_{DIoU}(\mathcal{M}, B_i) \geq \varepsilon \end{cases} \\ \mathcal{R}_{DIoU} = \frac{d^2}{c^2} \end{gathered} \tag{7}$$

where $\widetilde{s}_i$ is the updated confidence score, s_i is the initial confidence score, ε is the pre-set threshold, $\mathcal{M}$ is the box with the highest confidence score, B_i represents other overlapping boxes with $\mathcal{M}$, d is the Euclidean distance between the center points of the two boxes (represented by the red line segment in Figure 5), and c is the diagonal length of the smallest enclosing area (represented by the blue line segment in Figure 5).

3. Results

3.1. Comparison of Training Performance Between Two YOLOv5s Models

The hardware environment for this experiment includes a 13th Gen Intel(R) i5-13600KF CPU and an NVIDIA 3060Ti GPU, with the deep learning framework based on the Pytorch platform. The labeled spray rose dataset was input into both the original and improved YOLOv5s models, with all other training conditions kept consistent and the number of epochs set to 300. The evaluation metrics for the object detection models included precision (P), recall (R), mAP@0.5, and mAP@0.5:0.95 (as shown in Table 4). P and R are defined as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{8}$$

$$R = \frac{T_P}{T_P + F_N} \tag{9}$$

where F_N is the number of missed detections, F_P is the number of detection boxes with IoU less than the set threshold, and T_P is the number of detection boxes with IoU greater than the set threshold.

Table 4. Comparison of training results between the two YOLOv5s models.

Model	Precision (%)	Recall (%)	mAP@0.5%	mAP@0.5:0.95%
YOLOv5s_Ori	89.2	79.7	90.5	77.2
YOLOv5s [31]	86.1	74.4	81.2	71.0
YOLOv5 [30]	75.68	—	85.45	—
YOLOv6 [30]	82.47	—	82.35	—
YOLOv7 [30]	75.78	—	83.35	—
YOLOv8 [30]	93.79	—	97.13	—
YOLOv8 [31]	95.3	96.5	82.4	72.5
YOLOv8_imp [31]	97.4	95.4	83.1	72.5
YOLOv5s_Imp	99.4	99.7	99.5	90.8

Note: '—' in Table 4 means this parameter was not provided in the reference.

The mAP is calculated as the area under the P-R curve:

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

Although there are currently no reported applications of YOLO models for grading detection in spray rose cut flowers featuring many flowers on an individual stem, references [30,31] utilized YOLO models to grade roses featuring a single flower on an individual stem. In reference [30], the degree of openness of the flowers was categorized into five grades based on their maturity. Conversely, reference [31] divided the roses into four grades by considering both flower opening and leaf defects. Some results from these two papers are presented in Table 4 for comparative analysis within this study. For the YOLOv5s model employed in this research, the Precision, Recall, mAP@0.5%, and mAP@0.5:0.95% were recorded as 89.2, 79.7, 90.5, and 77.2, respectively; these figures slightly exceed those reported in reference [31], which were 86.1, 74.4, 81.2, and 71.0, respectively. The primary reason for this discrepancy was that only one factor, flower opening, was considered in this paper, whereas both flower opening and leaf defects were taken into account for grading roses in reference [31]. Furthermore, due to the incorporation of a maturity index in reference [30], the precision of the YOLOv5 model was noted at a lower value of 75.68 compared to those found in the other two studies. While both referenced studies indicated that the YOLOv8 model demonstrates higher accuracy for rose grading tasks, it was noteworthy that our improved YOLOv5s model achieved an impressive precision rate of 99.4%. Additionally, findings from reference [30] suggested that the performance of YOLOv7 was inferior to that of YOLOv5.

The training process for both YOLOv5s models is shown in Figure 6. During the initial phase of model training, there was no significant difference in precision and recall between the two models. However, after approximately 20 epochs, the improved model, benefiting from small object detection and the CBAM attention mechanism, began to establish a clear advantage. Around 150 epochs, the output of both network models stabilized. As shown in Table 4, the precision and recall of the improved YOLOv5s model increased by 10.2% and 20.0%, respectively. Figure 7 presents the confidence loss curves for both models during the training process. A lower confidence loss indicated a stronger ability of the model to detect targets accurately. It was evident from both the training and validation sets that the improved YOLOv5s model consistently outperformed the original model, primarily due to the WIoU loss function employed in the improved model.

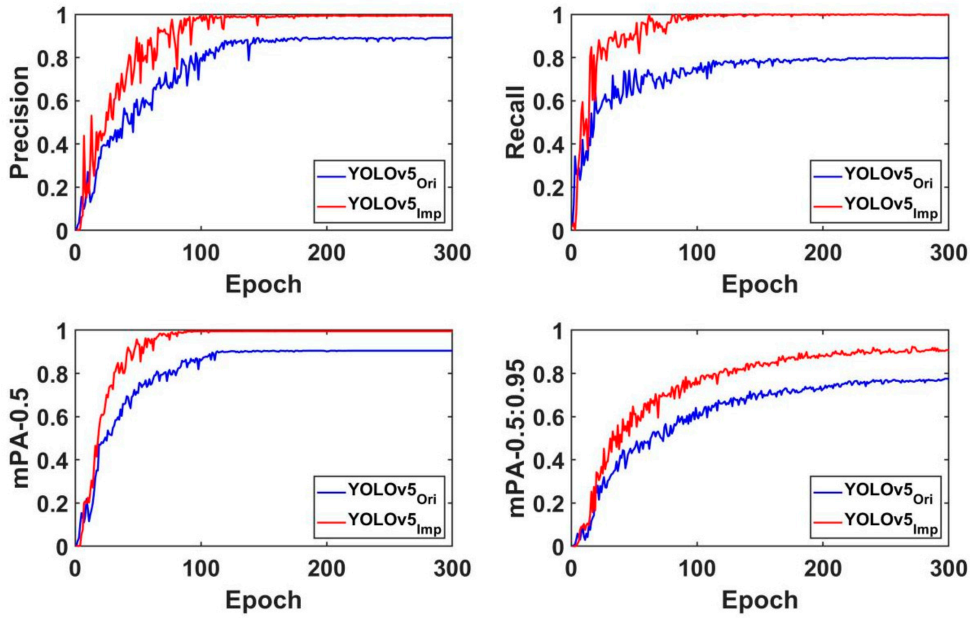


Figure 6. Performance comparison between the two YOLOv5s models.

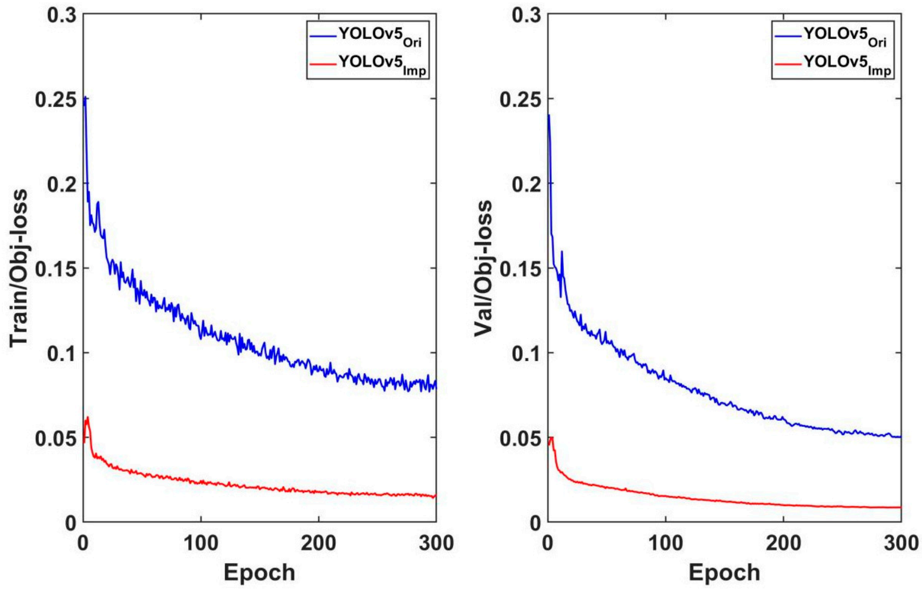


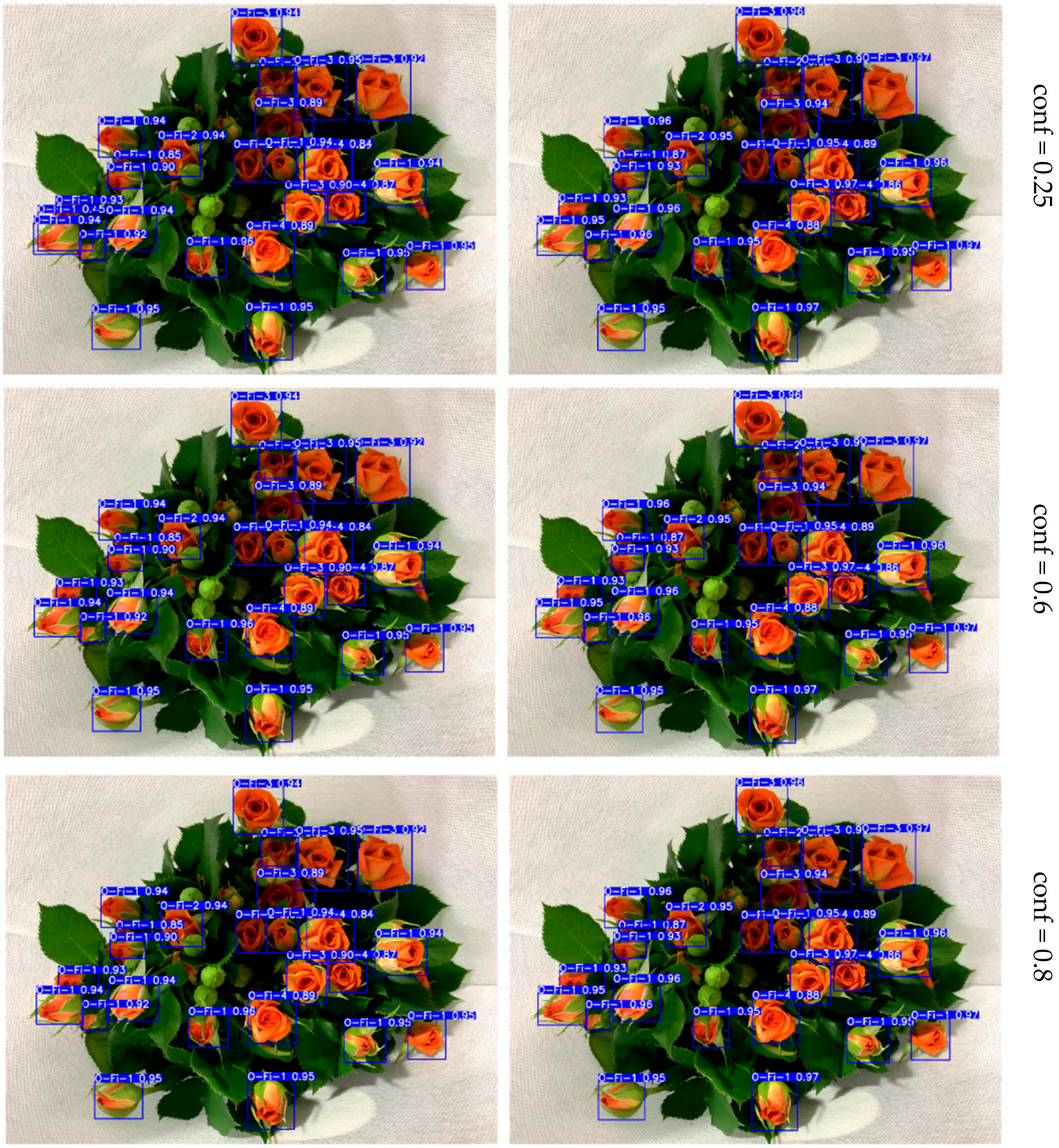
Figure 7. Confidence loss comparison between the two YOLOv5s models.

3.2. Comparison of Generalization Detection Capabilities Between Two YOLOv5s Models

To further assess the generalization capability of the models, two sets of images featuring packaged spray rose bouquets were randomly captured using a HUAWEI Mate 40 smartphone. One set comprised *Orange Bubbles*, while the other consisted of *Yellow Bubbles*. Given that these cut flower products are primarily sold outside the province, a higher proportion of roses with flowering indices 1 and 2 were included in the bouquets to accommodate extended transportation times, whereas roses with flowering index 5 were minimized. These two sets of images were subsequently input into both YOLOv5s models for detection purposes. With all other parameters held constant, confidence thresholds were established at 0.25, 0.6, and 0.8, respectively. The detection results were evaluated against manual annotations. For analytical purposes, two images each from Orange Bubbles and Yellow Bubbles bouquets were selected and labeled as Orange_*i* and Yellow_*i* (where $i = 1$, 2). Figures 8 and 9 present the detection results for Orange Bubbles bouquets; meanwhile, Figures 10 and 11 illustrate the detection outcomes for Yellow Bubbles bouquets. In each figure, the three sub-images in the left-hand column depict detection results obtained from the original YOLOv5s model under three different confidence thresholds: 0.25, 0.6, and 0.8, respectively. In contrast, those on the right-hand side represented detection results produced by an improved model. In these figures, "conf = 0.25" indicated that a confidence threshold was set to this value for analysis purposes. For comparative evaluation and analysis convenience, Tables 5 and 6 summarize the detection results pertaining to both types of rose bouquet images; blue text within these tables denoted findings derived from manual annotations.

(**a**) YOLOv5s_Ori (**b**) YOLOv5s_Imp

Figure 8. Detection results of Orange Bubbles bouquet (Orange_1) based on two YOLOv5s models.



(**a**) YOLOv5s_Ori (**b**) YOLOv5s_Imp

Figure 9. Detection results of Orange Bubbles bouquet (Orange_2) based on two YOLOv5s models.

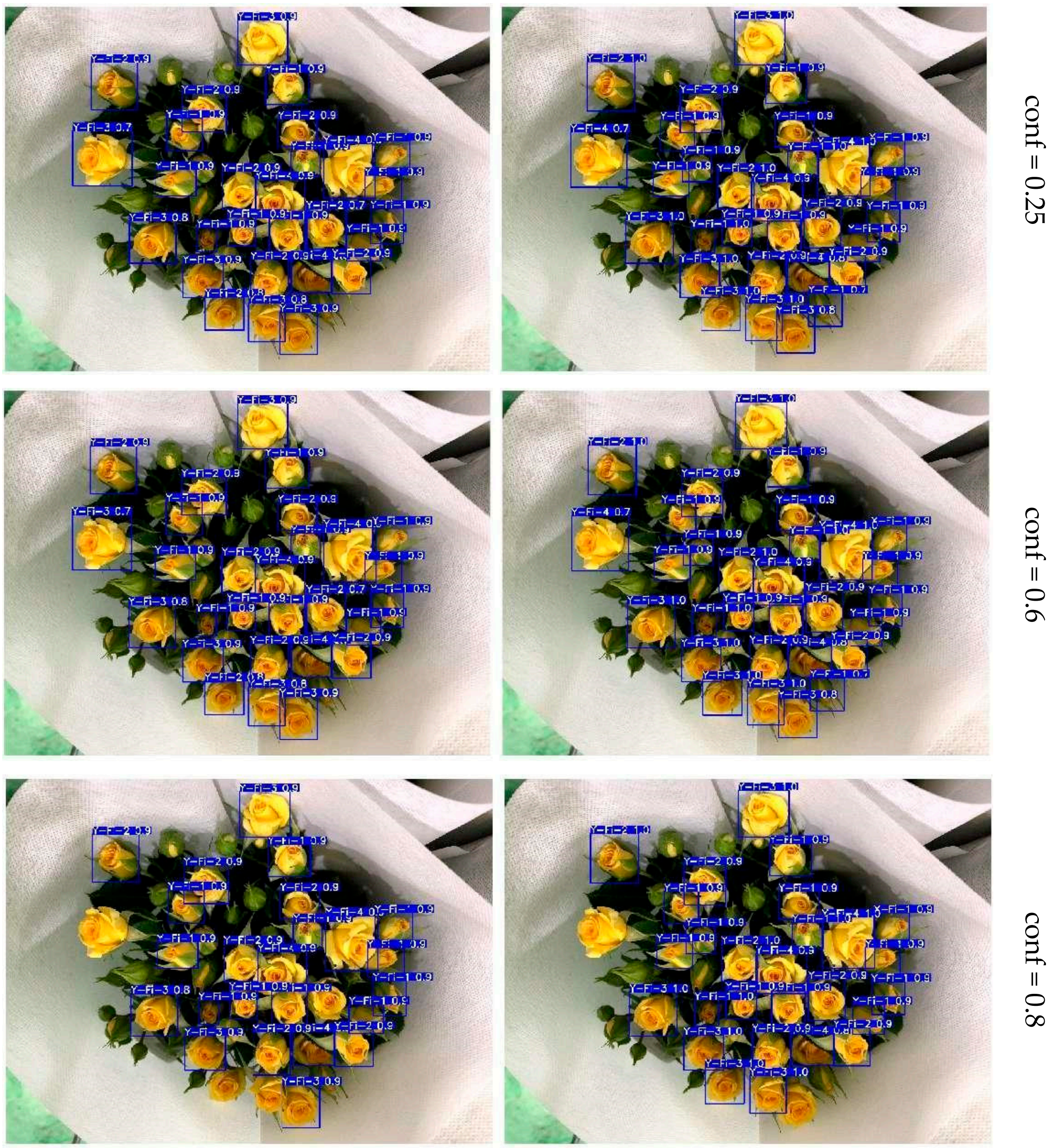
Table 5. Comparison of detection results on Orange Bubbles bouquets between two YOLOv5s models.

Test Rose	Conf	Model	O-Fi-1	O-Fi-2	O-Fi-3	O-Fi-4	Total
Orange_1	Manual Labeling		14	3	5	3	25
	0.25	YOLOv5s_Ori	15	3	7	3	28 + 2*
		YOLOv5s_Imp	15	3	5	3	26
	0.6	YOLOv5s_Ori	12	3	4	2	21
		YOLOv5s_Imp	14	3	5	3	25
	0.8	YOLOv5s_Ori	11	3	3	1	18
		YOLOv5s_Imp	12	3	5	3	23

Table 5. *Cont.*

Test Rose	Conf	Model	O-Fi-1	O-Fi-2	O-Fi-3	O-Fi-4	Total
Orange_2		Manual Labeling	11	7	4	2	24
	0.25	YOLOv5s_Ori	12	8	5	3	28
		YOLOv5s_Imp	11	9	5	3	28
	0.6	YOLOv5s_Ori	9	7	4	2	22
		YOLOv5s_Imp	11	7	4	2	24
	0.8	YOLOv5s_Ori	9	7	3	2	21
		YOLOv5s_Imp	10	7	4	2	23

Note: 2* indicates that 2 orange flowers are incorrectly identified as yellow ones of Y-Fi-5 (see Figure 12a).

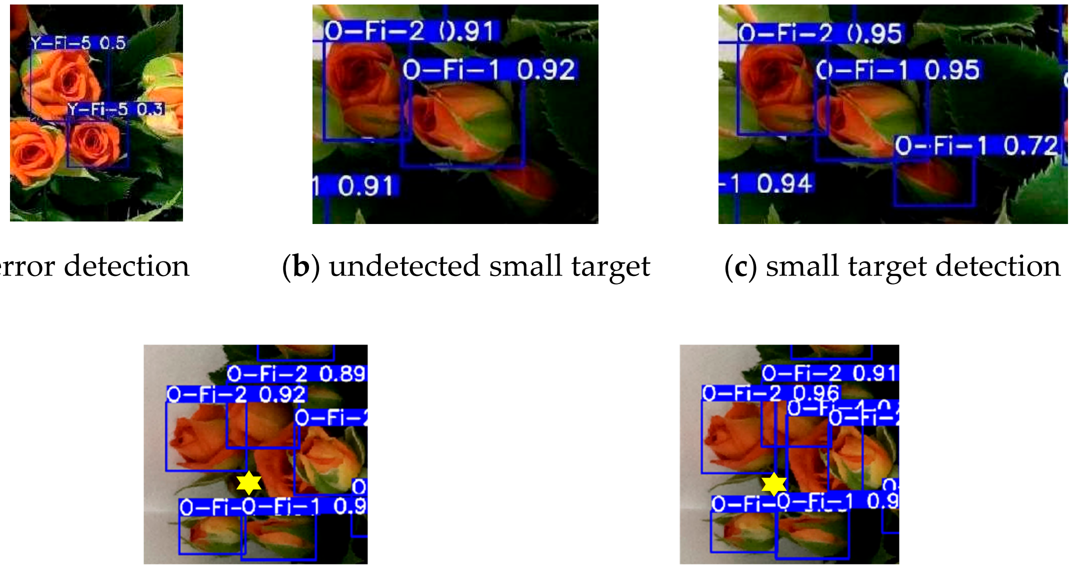


(**a**) YOLOv5s_Ori (**b**) YOLOv5s_Imp

Figure 10. Detection results of Yellow Bubbles bouquet (Yellow_1) based on two YOLOv5s models.

(**a**) YOLOv5s_Ori (**b**) YOLOv5s_Imp

Figure 11. Detection results of Yellow Bubbles bouquet (Yellow_2) based on two YOLOv5s models.



(**a**) error detection (**b**) undetected small target (**c**) small target detection

(**d**) undetected occlusion target (**e**) occlusion target detection

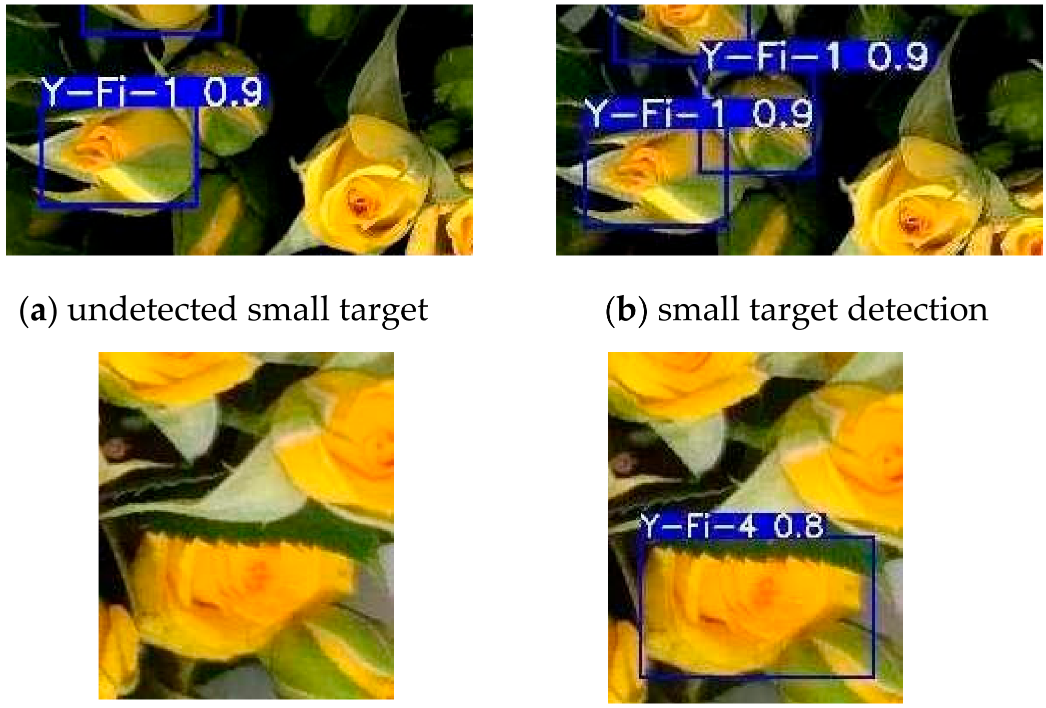
Figure 12. Partial enlarged view of the two Orange Bubbles bouquets.

Table 6. Comparison of detection results on Yellow Bubbles bouquets between two YOLOv5s models.

Test Rose	Conf	Model	Y-Fi-1	Y-Fi-2	Y-Fi-3	Y-Fi-4	Total
Yellow_1	Manual Labeling		13	7	6	4	30
	0.25	YOLOv5s_Ori	11	9	9	4	33
		YOLOv5s_Imp	14	6	8	5	33
	0.6	YOLOv5s_Ori	11	9	8	3	31
		YOLOv5s_Imp	13	7	6	4	30
	0.8	YOLOv5s_Ori	11	6	4	3	24
		YOLOv5s_Imp	12	7	6	4	29
Yellow_2	Manual Labeling		9	8	9	4	30
	0.25	YOLOv5s_Ori	11	12	10	3	36
		YOLOv5s_Imp	10	10	9	4	33
	0.6	YOLOv5s_Ori	9	11	9	3	32
		YOLOv5s_Imp	9	8	9	4	30
	0.8	YOLOv5s_Ori	9	9	7	2	27
		YOLOv5s_Imp	9	8	9	4	30

As shown in Tables 5 and 6, the detection results of the two models vary significantly with changes in the confidence threshold. When the confidence threshold was set to 0.25, both models detected more targets than those manually annotated, primarily because a lower confidence threshold can lead to multiple detections of the same target. Although there was no significant difference in the number of annotations between the two models, the original YOLOv5s model showed errors in annotation. For example, in the Orange_1 image, two orange roses with flowering index 4 were incorrectly labeled as yellow roses with flowering index 5 (Figure 12a).

When the confidence threshold was set to 0.6, the detection results of the improved YOLOv5s model were completely consistent with the manual annotations, while the original YOLOv5s model exhibited significant missed detections. For instance, in the Orange_2 image, the original YOLOv5s model could not find the small target (see Figure 12b for the Orange bouquet and Figure 13a for the Yellow one); however, the improved YOLOv5s model accurately detected small occluded roses due to its enhanced small object detection capability (see Figure 12c for the Orange bouquet and Figure 13b for the Yellow one).



(**a**) undetected small target (**b**) small target detection

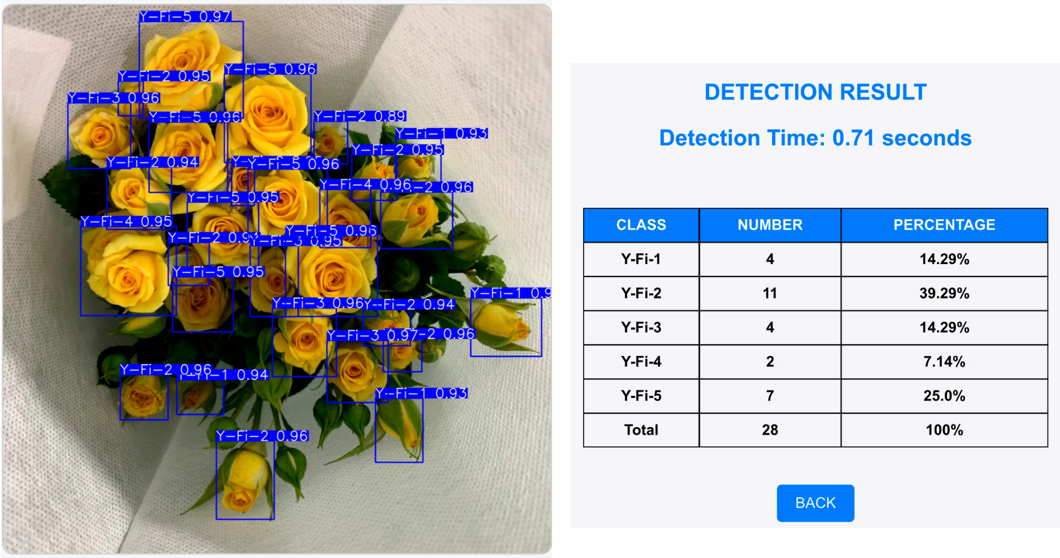
(**c**) undetected occlusion target (**d**) occlusion target detection

Figure 13. Partial enlarged view of the two Yellow Bubbles bouquets.

When the confidence threshold was increased to 0.8, the original YOLOv5s model detected significantly fewer roses than were manually annotated and failed to label relatively independent roses. In contrast, even under occlusion and blurred contours, the improved model was still able to accurately detect and label the rose. For instance, the occluded flower in Orange_1, as illustrated in Figure 8 and marked by a yellow asterisk on the partially enlarged images in Figure 12d,e, could not be detected by the original model. However, the improved version successfully identified and accurately labeled it. Additionally, for blurred targets, the enhanced model was also able to correctly detect and label them (see Figure 13d), which were undetectable by the original model (see Figure 13c).

The original model detected an average of 79% of the manually annotated roses in the Orange Bubbles bouquets and an average of 84% in the Yellow Bubbles bouquets. This discrepancy is mainly because the Orange Bubbles bouquets contained a significantly higher proportion of roses with flowering index 1 compared to the Yellow Bubbles bouquets. As shown in Tables 5 and 6, for images with a smaller proportion of roses with flowering index 1, the improved YOLOv5s model performed similarly to the manual annotations. However, when the proportion of roses with flowering index 1 was high, the improved model still exhibited occasional missed detections.

The above comparison results indicated that the improved YOLOv5s model achieved results consistent with manual annotations when the confidence threshold was set to 0.6. Even at a threshold of 0.8, the average detection rates for Orange Bubbles and Yellow Bubbles rose bouquets reached 95% and 99%, respectively, of the manually annotated quantities. Consequently, the improved YOLOv5s model was further deployed on a mobile phone (see Figure 14) to label the flowering spray rose cut flowers while providing the quantities and proportions of roses at different flowering index levels. The confidence threshold can be adjusted as needed to guide the grading of spray rose cut flowers. As depicted in Figure 14, the detection time for a bouquet of Yellow Bubbles roses was 0.71 s.



DETECTION RESULT

Detection Time: 0.71 seconds

CLASS	NUMBER	PERCENTAGE
Y-Fi-1	4	14.29%
Y-Fi-2	11	39.29%
Y-Fi-3	4	14.29%
Y-Fi-4	2	7.14%
Y-Fi-5	7	25.0%
Total	28	100%

BACK

Figure 14. Deployment of the improved YOLOv5s model on a mobile phone.

4. Discussion

This paper presented an enhanced YOLOv5s model for the classification and detection of flowering indices in spray rose cut flowers. The proposed detection algorithm improves upon the traditional YOLOv5s model by integrating the CBAM attention mechanism, thereby enhancing the model's capability to extract features specific to rose flowers. Furthermore, the CIoU loss function has been replaced with the WIoU loss function to better accommodate regression precision for bounding boxes of varying scales. To ensure accurate detection of occluded small targets within rose bouquets, a 160 $\times$ 160 small-scale feature output has also been incorporated. Additionally, the DIoU-NMS method was employed to optimize detection results effectively.

The training results on the same dataset showed that the improved YOLOv5s model achieved a 10.2% increase in precision and a 20.0% increase in recall. For randomly collected images of Orange Bubbles and Yellow Bubbles bouquets, the detection results of the improved YOLOv5s model were fully consistent with manual annotations at a confidence threshold of 0.6. Even when the confidence threshold was raised to 0.8, the detection accuracy for the flowering indices of both types of spray roses remained above 95%. Following the deployment of the enhanced YOLOv5s model on a mobile device, it successfully displayed flowering index labels, proportions, detection time, and other metrics for randomly captured rose bouquet images within just a few seconds.

Although the precision of the improved YOLOv5s model achieved 99.4% in this study, several shortcomings remained. Firstly, the training sample set exhibited a significant imbalance. The images were collected from a fresh-cut rose company to accommodate long-distance transportation needs; roses with high flowering indices were typically discarded. Consequently, most roses had a flowering index ranging from 1 to 3, while only 21.0% possessed a flowering index of 4 or 5. Secondly, the environment in which rose images were collected was relatively simplistic; in particular, varying lighting conditions were not taken into account. This factor also contributes significantly to the high detection precision observed in the improved YOLOv5s model presented in this paper. Additionally, only two varieties of spray roses were utilized for detection and grading within this study's application of the YOLOv5s model. For these reasons, we intend to further enhance our image dataset by incorporating a wider variety of samples. Additionally, we plan to utilize updated versions of YOLO models for the classification of spray rose flowers in our future work.

5. Conclusions

The enhanced YOLOv5s model presented in this paper offers a precise and efficient approach for assessing the openness of spray roses, thereby establishing a crucial foundation for the automated grading of spray rose cut flowers. However, it is important to note that the current model is limited to recognizing only two types of spray rose cut flowers within the orange and yellow color spectrum. In future work, the applicability of this network model can be broadened by increasing the sample size of spray roses across various color series based on their classification.

Author Contributions: J.L.: data curation, Methodology, validation, software, writing and review draft. M.L.: software, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded through a project funded by the College Students' Innovative Entrepreneurial Training Plan Program, grant number 202410307120Y.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on reasonable request.

Acknowledgments: We are grateful to Tonghai Jinhai Agricultural Technology Development Co., Ltd. for providing the image set of spray rose cut flowers.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Grishick, R.; Donahue, J.; Darrell, T.; Malik, J. Region based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [CrossRef] [PubMed]
2. Ren, S.; He, K.; Grishick, R.; Sun, J. Faster R-CNN: Towards real time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
3. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [CrossRef] [PubMed]
4. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
5. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
6. Mathew, M.P.; Mahesh, T.Y. Leaf-based disease detection in bell pepper plant using YOLO v5. *Signal Image Video Process.* **2022**, *16*, 841–847. [CrossRef]
7. Soeb, M.J.A.; Jubayer, M.F.; Tarin, T.A. Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Sci. Rep.* **2023**, *13*, 6078. [CrossRef]
8. Xue, Z.; Xu, R.; Bai, D.; Lin, H. YOLO-Tea: A Tea Disease Detection Model Improved by YOLOv5. *Forests* **2023**, *14*, 415. [CrossRef]
9. Tian, Y.; Wang, S.; Li, E.; Yang, G.; Liang, Z.; Tan, M. MD-YOLO: Multi-scale Dense YOLO for small target pest detection. *Comput. Electron. Agric.* **2023**, *213*, 108233. [CrossRef]
10. Wu, X.; Liang, J.; Yang, Y.; Li, Z.; Jia, X.; Pu, H.; Zhu, P. SAW-YOLO: A Multi-Scale YOLO for Small Target Citrus Pests Detection. *Agronomy* **2024**, *14*, 1571. [CrossRef]
11. Qin, K.; Zhang, J.; Hu, Y. Identification of Insect Pests on Soybean Leaves Based on SP-YOLO. *Agronomy* **2024**, *14*, 1586. [CrossRef]
12. Zhao, J.; Du, C.; Li, Y. YOLO-Granada: A lightweight attentioned Yolo for pomegranates fruit detection. *Sci. Rep.* **2024**, *14*, 16848. [CrossRef]
13. Lin, Y.; Huang, Z.; Liang, Y.; Liu, Y.; Jiang, W. AG-YOLO: A Rapid Citrus Fruit Detection Algorithm with Global Context Fusion. *Agriculture* **2024**, *14*, 114. [CrossRef]
14. Sun, Q.; Li, P.; He, C.; Song, Q.; Chen, J.; Kong, X.; Luo, Z. A Lightweight and High-Precision Passion Fruit YOLO Detection Model for Deployment in Embedded Devices. *Sensors* **2024**, *24*, 4942. [CrossRef] [PubMed]
15. Shi, Y.; Qing, S.; Zhao, L.; Wang, F.; Yuwen, X.; Qu, M. YOLO-Peach: A High-Performance Lightweight YOLOv8s-Based Model for Accurate Recognition and Enumeration of Peach Seedling Fruits. *Agronomy* **2024**, *14*, 1628. [CrossRef]
16. Wu, M.; Lin, H.; Shi, X.; Zhu, S.; Zheng, B. MTS-YOLO: A Multi-Task Lightweight and Efficient Model for Tomato Fruit Bunch Maturity and Stem Detection. *Horticulturae* **2024**, *10*, 1006. [CrossRef]
17. Steinmetz, V.; Delwiche, M.J.; Giles, D.K. Sorting cut roses with machine vision. *Trans. ASAE* **1994**, *37*, 1347–1353. [CrossRef]

18. Hiary, H.; Saadeh, H.; Saadeh, M. Flower classification using deep convolutional neural networks. *IET Comput. Vis.* **2018**, *12*, 855–862. [CrossRef]
19. Mete, B.R.; Ensari, T. Flower classification with deep CNN and machine learning algorithms. In Proceedings of the 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Piscataway, NJ, USA, 11–13 October 2019; pp. 1–5.
20. Anjani, I.A.; Pratiwi, Y.R.; Nurhuda, S.N.B. Implementation of Deep Learning Using Convolutional Neural Network Algorithm for Classification Rose Flower. *J. Phys. Conf. Ser.* **2021**, *1842*, 012002. [CrossRef]
21. Sun, X.; Li, Z.; Zhu, T.; Ni, C. Four-Dimension Deep Learning Method for Flower Quality Grading with Depth Information. *Electronics* **2021**, *10*, 2353. [CrossRef]
22. Wang, N.; Cao, H.; Huang, X.; Ding, M. Rapeseed Flower Counting Method Based on GhP2-YOLO and StrongSORT Algorithm. *Plants* **2024**, *13*, 2388. [CrossRef]
23. Oishi, Y.; Habaragamuwa, H.; Zhang, Y.; Sugiura, R.; Asano, K.; Akai, K.; Shibata, H.; Fujimoto, T. Automated abnormal potato plant detection system using deep learning models and portable video cameras. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102509. [CrossRef]
24. Bai, Y.; Yu, J.; Yang, S.; Ning, J. An improved YOLO algorithm for detecting flowers and fruits on strawberry seedlings. *Biosyst. Eng.* **2024**, *237*, 1–12. [CrossRef]
25. Qi, C.; Nyalala, I.; Chen, K. Detecting the Early Flowering Stage of Tea Chrysanthemum Using the F-YOLO Model. *Agronomy* **2021**, *11*, 834. [CrossRef]
26. Zhou, X.; Sun, G.; Xu, N.; Zhang, X.; Cai, J.; Yuan, Y.; Huang, Y. A Method of Modern Standardized Apple Orchard Flowering Monitoring Based on S-YOLO. *Agriculture* **2023**, *13*, 380. [CrossRef]
27. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [CrossRef]
28. Wang, J.; Gao, Z.; Zhang, Y.; Zhou, J.; Wu, J.; Li, P. Real-Time Detection and Location of Potted Flowers Based on a ZED Camera and a YOLO V4-Tiny Deep Learning Algorithm. *Horticulturae* **2022**, *8*, 21. [CrossRef]
29. Lyu, S.; Zhao, Y.; Liu, X.; Li, Z.; Wang, C.; Shen, J. Detection of Male and Female Litchi Flowers Using YOLO-HPFD Multi-Teacher Feature Distillation and FPGA-Embedded Platform. *Agronomy* **2023**, *13*, 987. [CrossRef]
30. Duan, Z.; Liu, W.; Zeng, S.; Zhu, C.; Chen, L.; Cui, W. Research on a Real-Time, High-Precision End-to-End Sorting System for Fresh-Cut Flowers. *Agriculture* **2024**, *14*, 1532. [CrossRef]
31. Zhang, Y.Y.; Bing, S.Y.; Ji, Y.H.; Yan, B.B.; Xu, J.P. Grading Method of Fresh Cut Rose Flowers Based on Improved YOLOv8s. *Smart Agric.* **2024**, *6*, 118–127. (In Chinese)
32. *GB/T 41201-2021*; Grade of Rose Cut Flower. State Administration for Market Regulation: Beijing, China, 2021.