Watermelon Rootstock Seedling Detection Based on Improved YOLOv8 Image Segmentation

Qingcang Yu*, Zihao Xu, Yi Zhu
School of Computer Science and Technology, Zhejiang Sci-tech University, Hangzhou 310018, China

***Abstract*—Automated grafting is an important means for modern agriculture to improve production efficiency and graft seedling quality, among which the use of visual systems to quickly segment target rootstock seedlings is the key technology to achieve automated grafting. This study aims to solve the problems of inaccurate image segmentation and slow detection speed in traditional rootstock seedling segmentation algorithms. To address these challenges, this study proposes a lightweight segmentation method based on an improved version of YOLOv8s-seg. The improved YOLOv8-seg introduces FasterNet as the backbone network and designs an RCAAM module to enhance feature extraction ability and lightweight model. The D-C2f module is improved to enhance feature fusion ability, achieving efficient and accurate segmentation of watermelon rootstock seedlings and improving grafting efficiency. This article designs a series of comparative experiments, comparing the improved version of YOLOv8-seg with classic models such as Unet, SOLO v2, Mask R-CNN, Deeplabv3+ on a test set containing watermelon rootstock seedlings, and evaluating the recognition performance and detection effect of the model. The experimental results show that the improved version of YOLOv8-seg outperforms other models in mAP coefficient index and can segment seedlings more accurately. This study provides reliable deep learning-based solution for the development of automatic grafting robots, which can effectively reduce labor costs and improve grafting efficiency, meeting the requirements of automated equipment for inference efficiency and hardware resources.**

Keywords*—Image segmentation; YOLOv8s-seg; lightweight; deep learning

I. Introduction

Grafting of watermelon rootstock seedlings is a key technology widely used in melon cultivation, which can improve the disease resistance, adaptability, and yield of watermelons through grafting, and is highly valued by agricultural producers [1]. In recent years, with the continuous growth of demand in the high-quality watermelon market, watermelon grafting technology has gradually become standardized and scaled up [2]. At present, most grafting operations use traditional manual methods, but manual grafting has problems such as low efficiency and high cost, which restrict the further promotion of grafting technology and the efficient development of the watermelon industry. In addition, due to the precise cutting and combination of the delicate characteristics of the seedlings during the grafting process, the skill requirements for operators are high, and even a slight carelessness may affect the survival rate of grafting. In order to achieve the large-scale and intelligent development of watermelon rootstock seedling grafting, the research and development of efficient and intelligent grafting equipment has become an important direction for promoting the modernization of the watermelon industry [3]. Automated grafting is an important means for modern agriculture to improve production efficiency and graft seedling quality, among which the use of visual systems to quickly segment target rootstock seedlings is one of the key technologies for achieving automated grafting. If the segmentation is not accurate, it will not only reduce the efficiency of grafting, but may also lead to failure, affecting the subsequent growth and survival rate of seedlings [4]. Therefore, utilizing advanced visual systems and image processing techniques for real-time segmentation and localization of target areas can effectively improve the accuracy and consistency of grafting operations, and reduce the cost and risk of manual intervention [5].

However, traditional methods for identifying rootstock leaves have significant limitations in practical applications, making it difficult to meet the precision and reliability requirements of automated grafting. Rootstock leaves usually have miniaturization characteristics. In addition, rootstock leaves are often mixed with surrounding branches, soil, or other plant leaves, and the background texture is complex and varied. In the process of leaf edge segmentation, traditional methods often have jagged and uneven segmentation boundaries, especially when facing diverse shapes or overlapping leaves, their performance is particularly inadequate. When the lighting conditions are uneven, the background is complex, or the leaf targets are small, traditional visual detection models are more prone to false positives and false negatives. These issues not only reduce the accuracy of identifying rootstock leaves, but may also lead to grafting failure or mechanical equipment misoperation, increasing the risks and costs of agricultural production.

Existing segmentation networks such as Unet, SOLO v2, Mask R-CNN, and Deeplab v3+ are simple and highly accurate in seedling segmentation tasks, they have several limitations that make them less suitable for the problem at hand. Specifically, these networks are difficult to perform shallow feature aggregation, resulting in poor segmentation performance on small and low contrast seedling leaves. In addition, their spatial perception ability is relatively weak, making it difficult to accurately depict the edges of overlapping or obscured seedlings. In addition, many of these models have high computational costs, which makes them perform poorly in real-time applications in actual agricultural production

environments. On the other hand, the YOLOv8 algorithm chosen in this article surpasses its predecessors in the YOLO series in terms of recognition accuracy, speed, and real-time performance. Its efficient feature extraction and multi-scale processing capabilities make it particularly suitable for seedling segmentation tasks. However, despite YOLOv8's strong performance on public datasets, there are still certain limitations in handling small object segmentation and complex background interference. To address these challenges, this paper proposes an improved YOLOv8 neural network architecture that enhances feature extraction, fusion mechanisms, and segmentation accuracy, making it more effective in rootstock seedling segmentation. By addressing the limitations of existing methods and leveraging the advantages of YOLOv8, the proposed method ensures high segmentation accuracy and real-time applicability, providing a powerful solution for automatic grafting. In the feature extraction stage, YOLOv8 utilizes its lightweight and powerful object detection capabilities as the backbone network to quickly segment the overall contour of the rootstock. In the feature extraction stage, FasterNet lightweight network is used to maintain high computational efficiency while achieving excellent performance in feature extraction. And integrate the RCAAM module into the backbone to alleviate the problem of high-frequency information loss in deep feature images. The D-C2f module was introduced in the feature fusion stage to enhance the learning ability of morphological features of different watermelon rootstock seedlings. By integrating these improvements, YOLOv8 demonstrates outstanding performance in tasks involving small and complex targets. It achieves higher segmentation accuracy, reduces false positives, and has stronger adaptability to different conditions, consolidating its position as the most advanced model in precision agriculture applications. The experimental results show that the neural network significantly improves the accuracy and real-time performance of rootstock seedling recognition tasks under complex backgrounds and varying lighting conditions. Compared with traditional models, this method exhibits stronger robustness and adaptability in detecting key parts of rootstock seedlings, providing reliable technical support for automated grafting equipment.

At the end of the introduction, the structure of this article is summarized as follows: Section III provides a detailed explanation of the improved YOLOv8 architecture and the modifications made to improve the accuracy of seedling recognition. The Section IV describes the experimental setup, including the dataset used and the evaluation metrics used to assess model performance. The Section V introduces the results and highlights the model proposed in this paper in comparison to other sub models. Finally, the Section VI summarizes the potential application exploration and future research directions of this model in real automatic grafting robots. This structure aims to guide readers through research and provide a clear understanding.

II. Related Work

Historically, grafting operations mainly relied on manual labor, which was not only time-consuming and labor-intensive, but also easily limited by workers' experience and technical level. In the process of leaf edge segmentation, traditional methods often have jagged and uneven segmentation boundaries, especially when facing diverse shapes or overlapping leaves, their performance is particularly inadequate [6]. When the lighting conditions are uneven, the background is complex, or the leaf targets are small, traditional visual detection models are more prone to false positives and false negatives [7]. Scholars have conducted in-depth research on the grafting process using advanced visual algorithms. In 2013, He et al. [8] proposed a method based on machine vision using ellipse fitting to restore seedling leaf surfaces and extract parameters for robot automatic grafting in order to improve the automation level of fruit and vegetable grafting robots. In 2015, Zhang et al. [9] proposed a comprehensive image processing algorithm to extract feature information of grafting seedlings for relevant vegetable grafting robots. The rapid target recognition technology achieved through visual systems can not only significantly improve grafting efficiency, but also reduce labor costs and intensity, while improving the quality and consistency of grafted seedlings. It is an important direction for promoting the intelligent and precise development of modern agriculture.

With the development of deep learning and computer vision technology, image-based seedling recognition methods have been widely studied and applied. Zuo et al. [10] proposed a crop seedling plant segmentation network model that integrates semantic and edge information of the target region in order to accurately segment crop seedlings in natural environments and achieve automatic measurement of seedling position and phenotype. The experimental results show that under the same network training parameters, the average cross merge rate and average recall rate obtained by testing the method proposed in this paper are 58.13% and 64.72%, respectively, which are better than the segmentation results corresponding to manually labeled samples; In addition, after adding 10% of outdoor seedling images to the training samples, the average pixel accuracy of this method on the outdoor test set can reach 90.54%, demonstrating good generalization ability. Image processing technology can be used for high-throughput collection and analysis of crop population phenotypes, which is of great significance for crop growth monitoring, seedling condition assessment, and cultivation management. However, existing methods rely on empirical segmentation thresholds, resulting in insufficient accuracy in extracting phenotypes. Li et al. [11] proposed a method for extracting phenotypes from aerial images of maize seedlings, using maize as an example. Explored an end-to-end segmentation network called PlantU-net, which uses a small amount of training data to achieve automatic segmentation of overhead images of maize seedling populations. Automatically extract morphological and color related phenotypes, including maize stem coverage, external radius, aspect ratio, and plant orientation plane angle. Ma et al. [12] proposed a method for processing greenhouse vegetable leaf disease symptom images in order to achieve robust segmentation, as uneven lighting and cluttered backgrounds are the most challenging problems in disease symptom image segmentation. The results show that the overall accuracy of the proposed method is 90.67%, indicating that the method can obtain robust segmentation of disease symptom images.

III. IMPROVED YOLOV8 MODEL

A. YOLO v8

The YOLO network has undergone various improvements to address the challenges brought by early versions, aiming to enhance its adaptability to specific tasks while maintaining a balance between detection speed and accuracy. The improved version YOLOv8-Seg adopts a modular design, including three main components: backbone, neck, and head. The backbone extracts basic features from the input image, the neck processes and integrates these features at multiple scales, and the head generates the final prediction, including object classification, bounding box coordinates, and segmentation masks.

The high accuracy of the YOLOv8 model is attributed to the replacement of the C3 module with the C2f module in the YOLOv5 backbone network and neck network. The C2f module first goes through a Conv, uses the chunk function to evenly split the out into two vectors, and then saves them to a list. The latter half is input into a Bottleneck Block, which contains n Bottlenecks. Each Bottleneck output is appended to the list. In the YOLOv8 model, the head part of the prediction head has undergone significant changes compared to the YOLOv5 model. It has been replaced from an anchor box based object detection algorithm to an anchor box free object detection algorithm, which has the advantages of fast convergence and improved regression performance. The decoupling head structure is adopted to separate the regression branch from the prediction branch, and the integral form representation method proposed in the Distribution Focal Loss strategy is used for the regression branch. The coordinates are transformed from a deterministic single value prediction to a distribution. Compared to using coupling heads, decoupling heads can effectively reduce the computational complexity of model segmentation of rootstock seedlings, which not only accelerates processing speed but also enhances the generalization ability and robustness of rootstock seedling recognition.

B. Improved YOLOv8 Model

In the process of rootstock seedling segmentation, the traditional method often appears jagged and unsmooth segmentation boundary, especially in the face of morphological diversification or leaf overlap, uneven lighting conditions, complex backgrounds, or small leaf targets, this paper proposes an improved YOLOv8 neural network architecture, which can be used to improve the performance of the neural network, for efficient and accurate segmentation of cotyledon parts of rootstock seedlings, the structure is shown in Fig. 1. In the feature extraction stage, YOLOv8 utilizes its lightweight and powerful target detection ability as the backbone network to quickly identify the overall contour of the rootstock. However, YOLOv8 has some limitations in the extraction of detailed features. To make up for this deficiency, the advanced feature extraction ability of FasterNet is integrated to capture finer features through deep convolution, especially in the localization of small target sites. At the same time, the RCAAM module is designed to non-uniformly weight the key features in the spatial and channel dimensions to highlight the useful information and suppress the interference of background noise. Combined with the task-aware classification and positioning module, the recognition accuracy and location accuracy of rootstock seedlings are further improved. In addition, the D-C2F module was improved to achieve effective fusion of multi-scale features through dynamic convolution, highlighting high-resolution features of key parts of rootstocks from coarse to fine, and improving the efficiency of rootstock management, it is used to improve the learning ability of different morphological rootstock seedling characteristics, and the deformable modeling is used to adapt to the target morphological change characteristics.

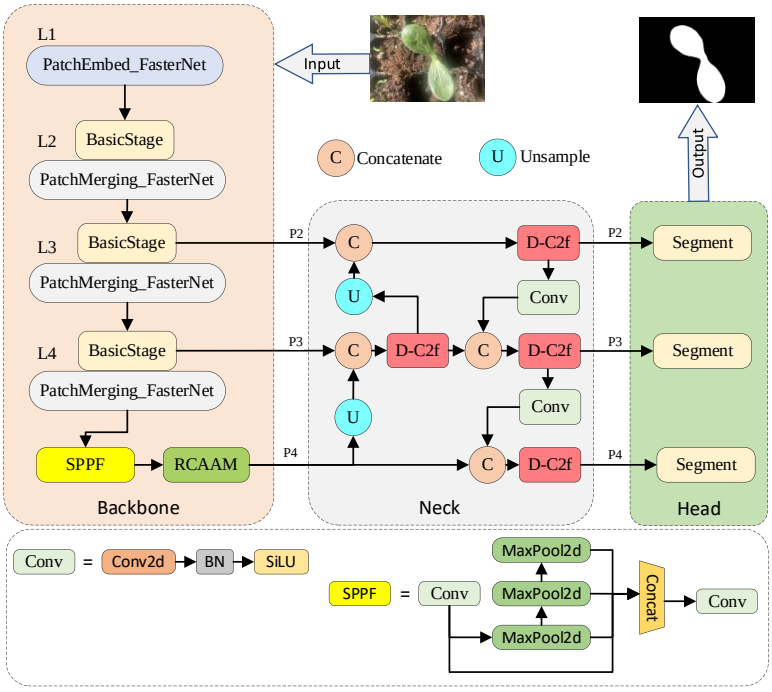

Fig. 1. Architecture of improved YOLOv8.

1) Based on FasterNet feature extraction: YOLOv8 uses DarkNet-53 as the backbone network, and its architecture consists of a series of convolutional layers and residual modules. Although DarkNet-53 performs well in general object detection tasks, it faces certain limitations in rootstock seedling segmentation, such as excessive redundant information, which limits training and inference speed, and insufficient segmentation performance for small target leaves and complex backgrounds. In addition, traditional backbone networks have certain bottlenecks in multi-scale information processing, making it difficult to effectively capture the global and local features of rootstock leaves. To address the aforementioned issues, CHEN et al. [13] proposed an efficient neural network called FasterNet, whose structure is shown in Fig. 2. Partial Convolution (PConv) is introduced in the article, which reduces redundant calculations and memory access by focusing on the features of specific regions, while significantly improving the efficiency of multi-scale feature extraction. This improvement enables FasterNet to achieve efficient model deployment on edge devices, while enhancing its ability to recognize small targets in complex scenes. In the task of rootstock seedling segmentation, combining FasterNet with YOLOv8 can effectively compensate for the shortcomings of DarkNet-53 in detail processing and feature extraction, providing new technical support for precise segmentation of rootstock leaves in complex agricultural scenes.

PConv only applies regular convolution to known regions in the input feature map, keeping other parts unchanged. The floating-point operands (FLOPs) of regular convolution and DWConv can be represented as:

$$F_{SC} = h \times w \times k^2 \times c^2 \tag{1}$$

$$F_{DWC} = h \times w \times k^2 \times c \tag{2}$$

In practical applications, PConv usually selects the first or last consecutive c_p channel, and requires that the input and output feature maps have the same channel. Therefore, the FLOPs of a PConv can be expressed as:

$$F_{PC} = h \times w \times c_p^2 \times k^2 \tag{3}$$

In the formula, cp=c/4, then the FLOPs of PConv are only 1/16 of those of regular convolution.

Its memory efficiency is significantly improved because PConv reduces memory access and data transmission by treating specific channels as representatives of the entire feature map, thereby accelerating computation speed without sacrificing accuracy. This paper replaces the backbone network of YOLOv8 with FasterNet to reduce redundancy and improve the overall computing speed, thus promoting efficient edge computing applications.

2) Residual channel adaptive attention module (RCAAM): To alleviate the problem of high-frequency information loss caused by the decrease in the number of convolutional channels in each layer, Wang et al. [14] proposed an adaptive attention module (AAM) to improve object detection in multi-scale scenes. Although this module improves the ability of feature extraction by adjusting the weight allocation of multi-scale features through adaptive average pooling, its operations mainly focus on low dimensional operations, resulting in an increase in shallow information redundancy and interference phenomena in feature images.

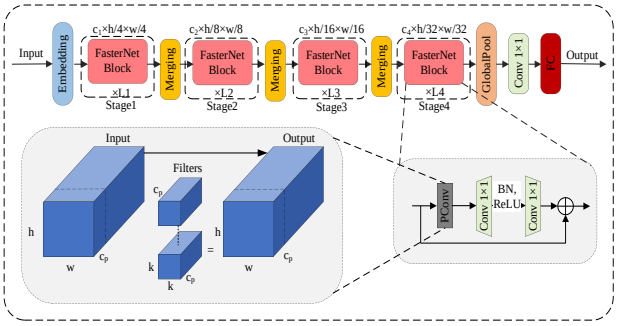

Fig. 2. Architecture of FasterNet.

To this end, this article has redesigned the AAM module and combined it with the residual channel attention network (RCAN) proposed by Zhang et al. [15], introducing super-resolution technology to enhance the detail representation and clarity of feature images. The core module in RCAN, residual group (RG), enriches feature encoding information through pixel addition strategy and enhances channel perception weights to recover more high-frequency information from the bottom layer, as shown in Fig. 3.

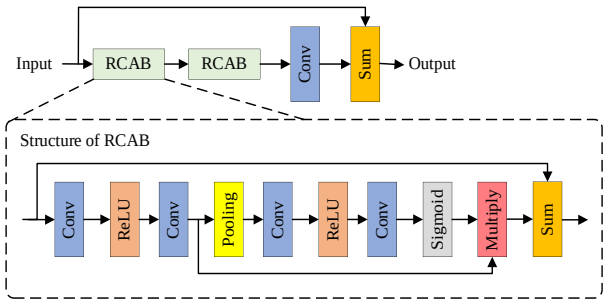

Fig. 3. Architecture of residual group.

This article proposes a new residual channel adaptive attention module (RCAAM) based on residual groups. RCAAM generates multiple sets of high-resolution feature maps in low resolution feature map reconstruction, which not only reduces redundant computation but also significantly restores shallow feature information of occluded targets.

In addition, this article introduces ECA module in RCAAM, which enhances contextual information correlation by weighted fusion of feature maps, thus more accurately segmenting rootstock seedlings in complex backgrounds. This article also compared multiple attention mechanisms, and the Seg loss function of the YOLOv8 model is shown in Fig. 4, with the model incorporating the ECA module performing the best in terms of loss function. The structure of the RCAAM module is shown in Fig. 5.

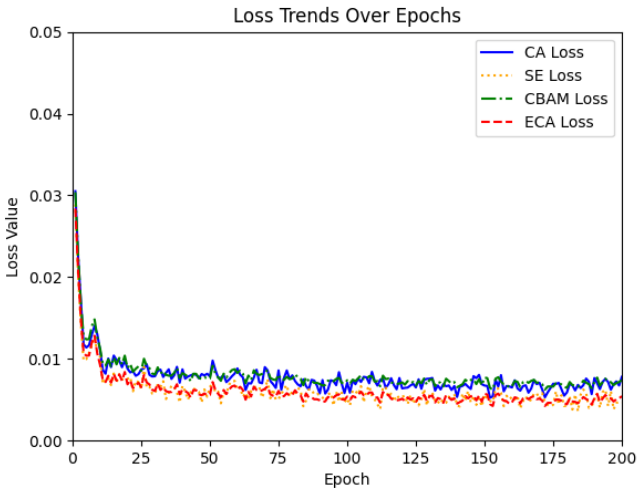

Fig. 4. Seg loss function curves for different attention mechanisms.

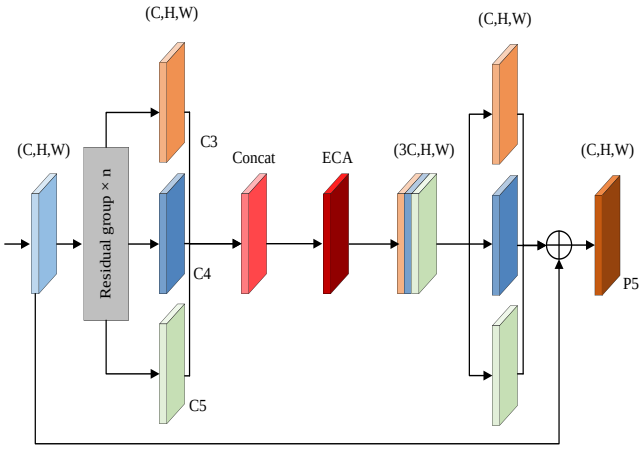

Fig. 5. The structure of the RCAAM module.

3) D-C2f Module: During the growth process of rootstock seedlings, their leaf shapes exhibit a high degree of diversity and are often accompanied by overlapping phenomena. In addition, due to the fact that rootstock seedlings usually grow in a plug environment, there are a large number of ridges and soil in the background, which further increases the difficulty of segmentation and recognition. Traditional standard convolutional neural networks (CNNs) use fixed convolution kernels for feature learning, which makes it difficult to fully capture the diverse morphological features of rootstock leaves, especially in cases of leaf overlap and blurred boundaries, leading to inaccurate feature extraction, false positives, and missed detections.

To address the above issues, this paper introduces deformable convolution (DConv) to enhance the model's ability to learn diverse rootstock leaf features. DConv adaptively adjusts the size and sampling position of the convolution kernel through deformable modeling, in order to better adapt to the changing characteristics of the target shape. Its structure is shown in Fig. 6.

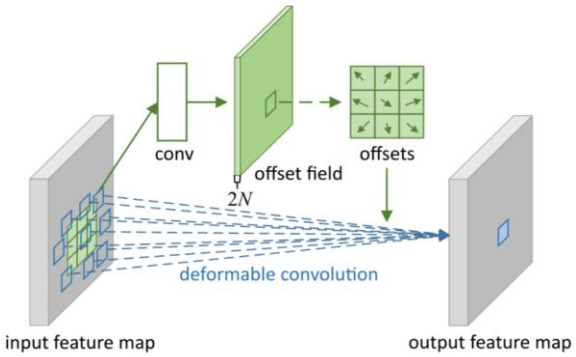

Fig. 6. The structure of the DConv module.

Specifically, DConv introduces offset in the convolutional receptive field and dynamically adjusts the position of each convolution kernel. For example, for a 3x3 convolution kernel, the regular sampling grid can be expanded by the offset, as shown:

$$R = \{\Delta p_n \mid n = 1,2,\cdots,n\} \tag{4}$$

By combining modulation variables, the model can automatically learn the offset and weight of each sampling point, enabling the sampling points to fit the target shape more accurately. For any position p, the output mapping Y is expressed as:

$$Y_{(p)} = \sum_{k=1}^{K} \omega_k \cdot x(p_k + \Delta p_k) \cdot \Delta m_k \tag{5}$$

Among them, ω_k and Δm_k respectively represent the weight and modulation variable of the kth position. Due to the fact that the offset Δp_k is usually in fractional form, bilinear interpolation is used for calculation, as shown in Eq. (6) to Eq. (8).

$$g(q_y, p_y) = \max(0, 1 - |a - b|) \tag{6}$$

$$G(q, p) = g(q_x, p_x) \cdot g(q_y, p_y) \tag{7}$$

$$X(p) = \sum_{q} G(q, p) \cdot g(q_y, p_y) \tag{8}$$

The application of DConv can dynamically adjust the sampling position and convolution kernel size according to the specific morphology of rootstock leaves. Firstly, by preprocessing the input feature map, offset and modulation variables are generated to calculate the offset direction of pixel points and obtain irregularly distributed sampling points. Subsequently, these sampling points are used to resample the feature map and combined with the convolution kernel to calculate the final convolution result.

To further improve the performance of the model, this paper redesigns the D-C2f module based on DConv convolution, as shown in Fig. 7. The D-C2f module can adaptively adjust the size and sampling position of the

convolution kernel based on the local features of the rootstock leaves by introducing learnable deformation parameters and offset weights. Compared with traditional methods, this module provides a larger receptive field range in the output features, greatly improving the model's ability to extract rootstock leaf features and model overlapping leaf spatial distribution in plug scenes, thereby effectively improving the segmentation and recognition performance of rootstock seedlings.

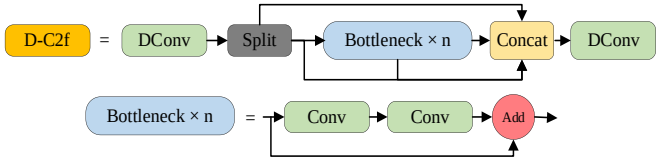

Fig. 7. The structure of the D-C2f module.

IV. Experiments And Analysis

A. Model Training and Testing Trials

1) Experimental data: The rootstock seedling task dataset of this study consists of images of watermelon seedlings in the cotyledon stage, constructed in two stages to ensure data diversity and representativeness. The first stage data was obtained from actual shooting at the seedling center, which captured 400 high-resolution images of watermelon seedlings. In the second stage, in order to unify and adapt the model training requirements, all images were preprocessed, cropped, and adjusted to a resolution of 640 × 640, while maintaining a 24 bit RGB format. After data augmentation and filtering, a total of 1600 images were generated, of which 1350 were divided into training and validation sets and randomly segmented in a 9:1 ratio. The remaining 250 images were used as the test set for independent evaluation of model performance, and the test set did not participate in model training. The dataset was annotated using the LabelMe tool and further converted to VOC dataset format. This dataset not only reflects the diversity of rootstock seedling characteristics, but also takes into account the plug scenes in real grafting environments, providing a solid foundation for training and evaluating segmentation and recognition models.

2) Comparison of segmentation network models: This study compared the performance of YOLOv8, Unet, SOLO v2, Mask R-CNN, and Deeplab V3+. Among them, SOLO v2 and Mask R-CNN belong to instance segmentation algorithms, while improved YOLOv8, Unet, and Deeplab v3+ belong to semantic segmentation models.

a) Unet: The Unet network structure is characterized by its clear U-shaped architecture, with symmetric encoders on the left and decoders on the right, which can effectively enhance the ability to extract feature map information. Unet has a low dependence on the number of images and only requires a small number of images to complete end-to-end training, making it very suitable for medical image segmentation. However, due to its relatively simple design, it is prone to inaccurate segmentation when dealing with complex backgrounds and small target tasks.

b) Mask R-CNN: Mask R-CNN is based on Faster R-CNN and is used to predict instance segmentation masks by adding a mask branch that runs in parallel with the classification and bounding box regression branches [16]. This method adopts a top-down detection approach, first detecting the regions of each instance, and then segmenting the instance masks within these regions. Detection based methods typically have high accuracy, but rely on precise bounding box detection, which places high demands on computational resources.

c) SOLO v2: Unlike Mask R-CNN, SOLO v2 transforms segmentation tasks into pixel classification problems, thereby eliminating the step of proposal generation [17]. The network consists of two branches: a category prediction branch for predicting the semantic category of the target, and a masking branch for predicting the instance mask of the target. This method reduces computational complexity and can improve the efficiency of instance segmentation to some extent, but its performance may be affected for scenes with complex backgrounds or overlapping targets.

d) Deeplab v3+: As the latest generation model in the Deeplab series, Deeplab v3+ adopts Deeplab v3 in its encoding structure and introduces a decoder to solve the problem of losing detailed information caused by directly upsampling feature maps in Deeplab v3, thus achieving higher performance in semantic segmentation tasks [18]. Deeplab v3+ has strong ability to recover detailed information, but it may still face certain challenges when dealing with small target tasks with complex backgrounds.

3) Testing trial setup: The hardware configuration is Intel i5-12490K CPU and Nvidia GeForce RTX 4060ti GPU. The experimental method of this paper is developed and implemented in Python 3.8 environment based on the deep learning framework PyTorch 2.2.2, using CUDA 12.4 for GPU acceleration. In the model training phase, the SGD optimizer is used to optimize the performance of the model, and the transfer learning strategy is used to initialize the model by loading the pre-trained model weight"yolov8s.pt" to accelerate the convergence speed of the model. The model training parameters in this paper are shown in Table I.

TABLE I. Experimental Model Parameters

Parameter Name	Value
Batch size	16
Epoch	200
Learning rate	0.01
Momentum factor	0.937
Image size	640 dpi×640 dpi

4) Evaluation metrics: In order to evaluate the segmentation performance of the improved YOLOv8 and the contrast model, this study uses the mean average Precision (mAP) as the main evaluation index, and combines Precision

and Recall to comprehensively analyze the performance of the model. Map is a commonly used metric in detection and segmentation tasks. Its calculation is based on the matching of the predicted results and the real labels, which can comprehensively measure the performance of the model in detection accuracy and coverage. See Eq. (9) and Eq. (10) for calculations of Precision and Recall.

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

In a single image, the calculation of mAP involves multiple target categories or instances. Firstly, calculate the accuracy and recall curves for each category: for a certain category, sort the model's predictions for that category in descending order of confidence, and gradually calculate the accuracy and recall at different thresholds. Secondly, calculate the average precision (AP): By integrating the precision and recall curves, obtain the AP value for that category. The higher the AP value, the stronger the detection ability of the model in that category. The AP formula is as follows:

$$AP = \int_0^1 P(R)dR \tag{11}$$

Finally, calculate the average precision mean (mAP) for all categories: take the average of the AP values for all categories, which is mAP, using the following formula:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{12}$$

where, N is the number of categories, and AP_i is the average accuracy of the i-th category.

In the segmentation task, the mAP calculation is usually based on the Intersection over Union (IoU) threshold setting. For example, when the threshold of IoU is set to 0.5, the ratio of the overlapping area representing the predicted result and the real label region to the joint area needs to be greater than 50% to be considered a correct detection. mAP@0.5 is a commonly used evaluation metric, in addition, the mAP can also be calculated at a higher IoU threshold to examine the sensitivity of the model to the target boundary.

5) Analysis of model training: In the training process of rootstock seedling segmentation algorithm, loss function plays an important role in evaluating the performance of the model. The smaller the loss value, the better the performance of the model and the more significant the optimization effect. As shown in Fig. 8, the training process of the improved algorithm generates a loss function curve, which intuitively reflects the obvious downward trend of the loss value with the number of iterations, indicating that the model performance is continuously optimized.

Among them, the decrease of box loss is the most significant, which reflects the improvement of target positioning accuracy. After approximately 75 training rounds, the Obj loss and seg loss tended to stabilize.

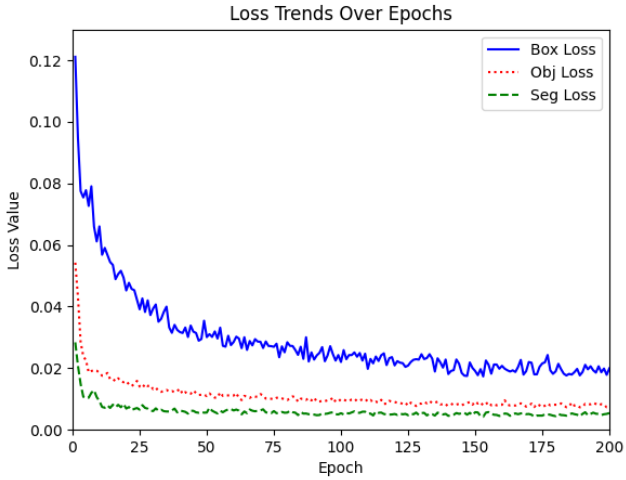

Fig. 8. Curve of model loss function.

The training process of YOLOv8 network before and after improvement is shown in Fig. 9. During the training process, the mAP of the model went through three stages: within the initial 25 training rounds, the mAP value rose rapidly, indicating that the model was in a rapid fitting stage; subsequently, between the 25th and 75th training rounds, the mAP value increased rapidly, indicating that the model was in a rapid fitting stage, the mAP value fluctuated greatly; from the 75 th to 200 th training round, the mAP curve tended to be stable and convergent with the change of learning rate gradually decreasing. With the increase of training times, the mAP value gradually converges, indicating that the model tends to be stable after further training.

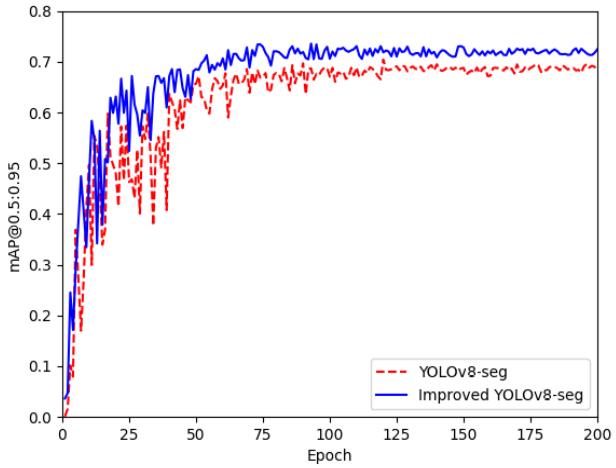

Fig. 9. Comparison of mAP curve changes with epoch variation.

B. Experimental Results

1) Ablation experiment: The purpose of ablation experiments is to explore the specific impact on model performance when certain parts of the network are removed or

modified. This article takes YOLOv8 as the baseline model, improves it in three aspects, designs an improved YOLOv8 model, and conducts five ablation experiments on the improved model. Among them, Improved Model 1 represents replacing the original backbone network with a FasterNet network, Improved Model 2 represents replacing the C2f module in the neck with a D-C2f module, and Improved Model 3 represents introducing the RCAAM module into the backbone network. The detailed experimental results are shown in TABLE II. I.

TABLE II. PERFORMANCE COMPARISON OF EACH IMPROVED MODEL

Model	FasterNet	D-C2f	RCAAM	P/%	R/%	mAP@0.5/%	Parameters/10^6	FLOPs/10^9
YOLOv8s	-	-	-	91.2	90.3	92.7	11.1	28.8
Improved model 1	√	-	-	91.4	93.6	95.3	6.0	11.8
Improved model 2	-	√	-	86.4	92.0	95.7	10.3	25.6
Improved model 3	-	-	√	95.6	91.2	94.4	7.8	21.8
Improved YOLOv8	√	√	√	95.4	95.0	96.6	3.2	10.6

According to the results in Table II, Improved Model 1 replaces the backbone network with a lightweight Faster Net network, mAP@0.5 Value increased by 2.6%, parameter and computation reduced by 5.1M and 17G. Indicating that lightweight FasterNet networks can reduce redundant information, optimize model parameter count and computational complexity, and improve detection speed. The FasterNet network adopts PConv to better extract multi-scale information of rootstock seedlings and cotyledons, improve detection ability and compress model volume, enhance the network's feature extraction ability, and reduce parameter counting.

Improved model 2 replaces the C2f module on the neck with a lighter D-C2f module, reducing the parameter and computational complexity by 0.8M and 3.2G compared to the baseline model, mAP@0.5. The value has increased by 3%. The D-C2f module can adaptively adjust the size and sampling position of the convolution kernel based on the local features of the rootstock leaves by introducing learnable deformation parameters and offset weights. This provides a larger receptive field range in the output features, effectively improving the segmentation and recognition performance of rootstock seedlings. Better save computational costs and improve training speed.

Improved model 3 then introduced the RCAAM module into the original backbone network, mAP@0.5 Compared to the baseline model, the value increased by 1.7%, while the parameter and computational complexity decreased by 3.3M and 7G, respectively. Compared with traditional pooling methods, RCAAM not only improves feature extraction efficiency, enhances contextual information correlation, but also suppresses useless information interference, thus more accurately segmenting rootstock leaves.

Finally, three modules were added simultaneously, namely the improved YOLOv8 algorithm proposed in this article, which increased mAP values by 3.9%, reduced parameter and computational complexity to only 28.8% and 36.8% of the baseline model, and achieved a recall rate of 95.0%. The ablation experiment results have verified the rationality and superiority of the algorithm proposed in this paper in terms of detection accuracy, speed, and lightweight.

2) Comparison of evaluation metrics between improved YOLOv8 and other models: In order to evaluate the performance of the improved YOLOv8 model in rootstock seedling segmentation, this study compared and analyzed the segmentation abilities of different models in the test set, focusing on their performance in handling significantly different backgrounds and complex low contrast scenes. To evaluate in detail the segmentation performance of the improved rootstock seedling recognition model, this study used mAP@0.5 Compare its performance with Unet, SOLO v2, Mask R-CNN, and Deeplab v3+ on the test set. The test set consists of 250 images, including two distinct types of seedlings, providing a diverse basis for performance evaluation. The results of evaluation metrics comparison are shown in TABLE III. .

TABLE III. COMPARISON BETWEEN MAINSTREAM SEGMENT ALGORITHMS AND THE PROPOSED METHOD

Model	P/%	R/%	mAP@0.5/%	Parameters/10^6	FLOPs/10^9
Mask-RCNN	87.3	88.4	92.3	44	44.7
SOLO v2	89.5	92.7	93.9	37.2	41
Deeplab v3+	84.8	85.0	88.9	41.2	12.6
Unet	91.3	93.3	94.7	31	8.3
Improved YOLOv8	95.4	95.0	96.6	3.2	10.6

The research results indicate that Mask R-CNN and Deeplab v3+ mAP@0.5 The indicators are significantly lower than the improved YOLOv8 model and Unet. This low performance reflects that these two models have difficulty accurately segmenting small seedlings under testing conditions. In contrast, improving the recognition model of rootstock seedlings mAP@0.5 The value is slightly higher than Unet, indicating its optimal performance among the four models. Specifically, the improved YOLOv8 model mAP@0.5 It reached 96.6%, indicating its significant advantage in the accuracy of rootstock seedling image segmentation and detection.

These results highlight the effectiveness of the improved model in segmenting rootstock seedlings and cotyledons under complex environmental conditions, and its stability and

robustness make it have important application potential in grafting management and production processes.

When segmenting rootstock seedling images with significant contrast between background and rootstock seedling cotyledons, the improved YOLOv8 model was compared with four other models (Unet, SOLO v2, Mask R-CNN, Deeplab v3+), and the results are shown in Fig. 10. The results indicate that Deeplab v3+ is relatively less effective than other models in segmenting small or edge blurred leaves. The segmentation performance of SOLO v2 and Mask R-CNN is superior to Deeplab v3+, but the computational cost is high and the performance is still insufficient when detecting small leaves or low contrast targets, which can easily miss some key leaf regions. The Unet model may encounter problems of over segmentation or under segmentation when segmenting overlapping or blurred boundary rootstock leaves, especially for leaves with complex shapes. The improved YOLOv8 model performs well in rootstock seedling segmentation tasks. Compared with other models, the improved YOLOv8 effectively enhances segmentation performance through stronger feature extraction ability and optimized feature fusion mechanism, and has excellent robustness and adaptability. The comparative experimental results show that the improved YOLOv8 model is suitable for precise segmentation of rootstock seedlings in complex environments, providing efficient and reliable technical support for automatic grafting in smart agriculture.

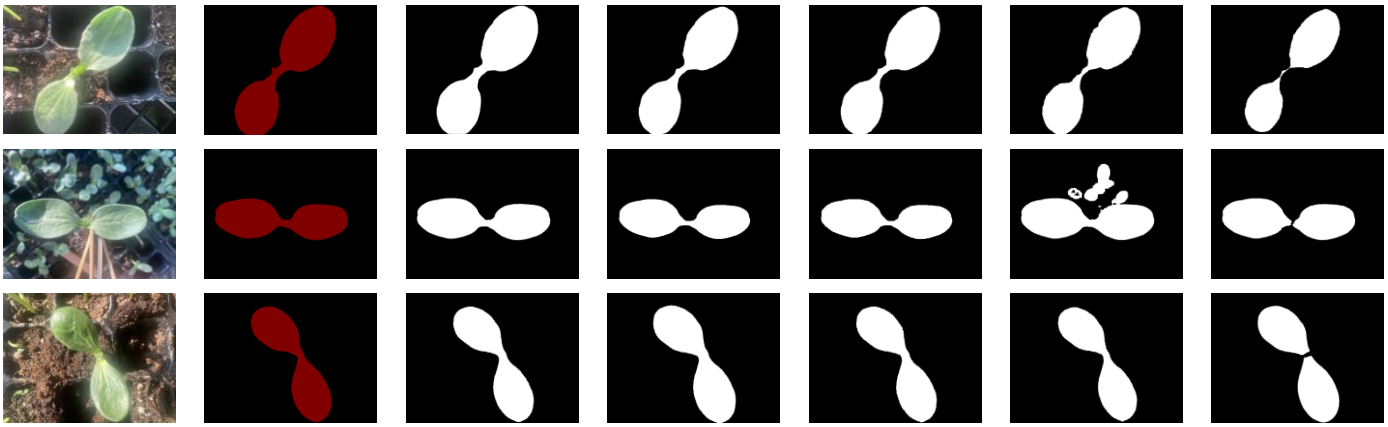

Fig. 10. Comparison of mAP curve changes with epoch variation.

V. Discussion

The improved YOLOv8 rootstock seedling recognition model outperforms the original YOLOv8 model in terms of recognition performance. In practical complex environments, the original model performs poorly in situations where lighting is uneven or the background is similar to the characteristics of rootstock seedlings, leading to inaccurate segmentation. The improved model can accurately segment rootstock seedlings with higher recognition rate, especially in complex environments, demonstrating stronger adaptability. At the same time, it effectively solves the problem of low accuracy in detecting small or distant rootstock seedlings in the original model.

This performance improvement does not come from the improvement of a single method, but from the comprehensive enhancement of feature extraction and feature fusion capabilities. By improving the network structure, optimizing the feature processing flow, and introducing more advanced feature fusion strategies, the robustness and segmentation accuracy of the model in complex environments have been significantly improved, better adapting to diverse hardware deployment requirements, providing more efficient and intelligent support for modern agricultural production, and promoting the implementation of precision agriculture. In order to ensure the efficient application of the rootstock seedling recognition system in practical agricultural production, it is possible to consider removing unimportant connections or neurons from the model in the future. Pruning technology can significantly reduce the size and inference time of the model while maintaining its recognition accuracy. This method is particularly suitable for devices with limited computing resources, making the rootstock seedling recognition system more efficient and meeting the real-time detection needs of resource constrained platforms. In large-scale plug seedling grafting, the segmentation model of the pruned rootstock seedlings can achieve real-time detection on unmanned aerial vehicles or vehicle platforms, improving the efficiency and accuracy of grafting.

VI. Conclusion

In order to solve the problems of inaccurate edge segmentation and low detection efficiency of traditional detection algorithms, this paper proposes a rootstock seedling detection method based on improved YOLOv8. By suppressing invalid features in high-order and low-order feature fusion and enhancing the ability of the model to extract rootstock seedling features, this method can effectively realize the recognition of narrow and small seedlings and large seedlings.

In this paper, mAP@0.5 is used as the evaluation index to compare the performance of the improved YOLOv8 and the basic UNET model in the segmentation of watermelon rootstock seedlings. The results show that the improved method is superior to the basic UNET algorithm in both quantitative and qualitative evaluation. In the aspect of model performance, the model parameters and FLOPs were used as evaluation criteria. Although the improved YOLOv8 outperforms Unet in recognition performance, its FLOPs are

2.3 g higher than Unet, showing a good balance between performance and recognition accuracy. Compared with the other three classical segmentation networks, the results show that the MAP@0.5 score of the improved YOLOv8 is 0.8403, which is better than the classical models such as SOLO v2, Mask R-CNN and Deeplab v3+. Compared with other models, the improved YOLOv8 model has the highest performance in identifying rootstock seedlings, and can accurately extract the characteristic information of seedlings.

The improved model has important application potential in watermelon rootstock grafting. Its ability to accurately segment the characteristics of small and irregular watermelon rootstock seedlings provides an important guarantee for improving the efficiency and accuracy of agricultural production. By integrating the improved model into the automatic grafting robot, the time and labor cost required for traditional manual grafting can be significantly reduced.

Although the improved YOLOv8 model performs well, there are still areas that need improvement in the future. Firstly, reducing computational complexity while maintaining high detection accuracy remains the main research direction. In future research, optimizing the model structure or using a more lightweight network architecture can be considered. Secondly, the actual agricultural production environment is more complex, with different lighting conditions, different seedling growth conditions, and different seedling varieties. Future work should improve the robustness of models in different environments. In addition, combining this model with edge computing equipment can enhance its practical applicability in precision agriculture.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (No.51375460).

REFERENCES

[1] Kyriacou M C, Rouphael Y, Colla G, et al. Vegetable grafting: The implications of a growing agronomic imperative for vegetable fruit quality and nutritive value[J]. Frontiers in plant science, 2017, 8: 741.

[2] Kumar P, Rouphael Y, Cardarelli M, et al. Vegetable grafting as a tool to improve drought resistance and water use efficiency[J]. Frontiers in plant science, 2017, 8: 1130.

[3] Lee J M, Kubota C, Tsao S J, et al. Current status of vegetable grafting: Diffusion, grafting techniques, automation[J]. Scientia Horticulturae, 2010, 127(2): 93-105.

[4] Maurya D, Pandey A K, Kumar V, et al. Grafting techniques in vegetable crops: A review[J]. International Journal of Chemical Studies, 2019, 7(2): 1664-1672.

[5] Gaion L A, Braz L T, Carvalho R F. Grafting in vegetable crops: A great technique for agriculture[J]. International Journal of Vegetable Science, 2018, 24(1): 85-102.

[6] Hétroy-Wheeler F, Casella E, Boltcheva D. Segmentation of tree seedling point clouds into elementary units[J]. International Journal of Remote Sensing, 2016, 37(13): 2881-2907.

[7] Scharr H, Minervini M, French A P, et al. Leaf segmentation in plant phenotyping: a collation study[J]. Machine vision and applications, 2016, 27: 585-606.

[8] He L, Cai L, Wu C. Vision-based parameters extraction of seedlings for grafting robot[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(24): 190-195.

[9] Zhang L, He H, Wu C. Vision method for measuring grafted seedling properties of vegetable grafted robot[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(9): 32-38.

[10] Zuo X, Lin H, Wang D, et al. A method of crop seedling plant segmentation on edge information fusion model[J]. IEEE Access, 2022, 10: 95281-95293.

[11] Li Y, Wen W, Guo X, et al. High-throughput phenotyping analysis of maize at the seedling stage using end-to-end segmentation network[J]. PLoS One, 2021, 16(1): e0241528.

[12] Ma J, Du K, Zhang L, et al. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing[J]. Computers and Electronics in Agriculture, 2017, 142: 110-117.

[13] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.

[14] Wang J, Chen Y, Dong Z, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection[J]. Neural Computing and Applications, 2023, 35(10): 7853-7865.

[15] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 286-301.

[16] Shen L, Su J, Huang R, et al. Fusing attention mechanism with Mask R-CNN for instance segmentation of grape cluster in the field[J]. Frontiers in plant science, 2022, 13: 934450.

[17] Zhou R J, Zheng L M, Ren C L, et al. Image Segmentation Algorithm in Complex Environment Based on Improved SOLOV2[C]//2023 12th International Conference of Information and Communication Technology (ICTech). IEEE, 2023: 581-585.

[18] Yang T, Zhou S, Xu A, et al. An approach for plant leaf image segmentation based on YOLOV8 and the improved DEEPLABV3+[J]. Plants, 2023, 12(19): 3438.