Identification and Segmentation of Camellia Oleifera Branches to be Pruned Based on Improved YOLOv8-seg

Pinglu Chen, Feng Huang, Jing Xu, and Muhua Liu

***Abstract*—An improved recognition and segmentation model for pruning branches of Camellia oleifera trees based on YOLOv8n-seg was proposed, aiming to address the challenges posed by dense interleaving of branches, severe occlusion, and complex shape characteristics. The SPPF module was replaced with the SPPFCSPC module, DCNv2 was introduced into the C2f module, and Wise-IoU v3 loss function replaced CIoU loss function. The experimental findings reveal that the enhanced YOLOv8n-seg model exhibits a remarkably compact size of merely 9.56MB, rendering it exceptionally well-suited for deployment. The recognition mean Average Precision (box-mAP) and the mask mean Average Precision (mask-mAP) on the dataset of Camellia oleifera branches necessitating pruning stand at an impressive 97.9% and 96%, respectively, surpassing the original model by 1.1% and 1.2%. This advancement effectively mitigates challenges associated with intricate shape features and occlusion in Camellia oleifera branches requiring pruning. Furthermore, the inference time for a single image is a swift 39.5ms.**

***Index Terms*—YOLOv8, Camellia oleifera, Instance segmentation, Pruning**

I. Introduction

CAMELLIA oleifera is one of the characteristic economic crops in China, belonging to the category of evergreen crops that thrive throughout the year. Its fruit possesses abundant oil content and holds significant economic value [1]. Pruning operation directly impact economic benefits of Camellia oleifera management and cultivation [2]-[3]. Currently, the pruning of Camellia oleifera relies on manual assistance with simple pruning knives, which is plagued by issues such as high labor intensity, low efficiency, and imprecise identification of branches requiring pruning. Developing a robotic system for Camellia oleifera pruning presents an effective solution to address these challenges. The accurate and rapid recognition of the branches requiring pruning by machines as one of the primary obstacles must be overcome during the development of pruning robot for Camellia oleifera. Compared to deciduous fruit trees, the lush branches and evergreen leaves of Camellia oleifera trees during the pruning process poses greater challenges for machine recognition of the branches to be pruned.

Although research on the machine recognition of Camellia oleifera tree branches is currently limited, recent studies have witnessed an upsurge in machine recognition of fruit tree branches due to its pivotal role in the advancement of pruning and picking robots for fruit trees. The skeleton of grape branches was extracted by Jia et al. [4] through a comparison of various thinning algorithms, based on the acquisition of binary images. Among these algorithms, the Rosenfeld thinning algorithm was selected for its superior ability to maintain skeleton connectivity. Huang et al. [5] proposed a method for image recognition and frame extraction of loquat branches. Firstly, the segmentation threshold was determined through brightness conversion to achieve accurate segmentation, and then the feature image was extracted by using the branch feature of loquat branches. Under the example verification, the correct recognition rate of loquat branch image was 91.2%. Amatya et al. [6] developed a machine vision system for segmentation and detection of cherry branches and leaves. Firstly, an image segmentation methodology was devised to discern the visible segments of the branches, followed by the application of a Bayesian classifier to categorize the image pixels into branches, cherries, foliage, and other pertinent classifications. The accuracy of branch recognition under this method was 89.6%. Zhang et al. [7] proposed an improved AlexNet network for apple branch detection and the average recall and accuracy of this method were 92% and 86%, respectively. Zheng et al. [8] proposed a lightweight Attention Ghost-HRNet (AGHRNet) network model based on deep learning to segment the trunk and branches of jujube trees, and the mIoU and mPA of the method were 77.79% and 89.46%, respectively.

The aforementioned research is based on the recognition of fruit tree branches using two-dimensional images. In order to further acquire the positional information of the targeted branches, certain scholars have conducted three-dimensional point cloud recognition of fruit tree branches. Karkee et al. [9] proposed a machine vision system, employing a three-dimensional camera to capture three-dimensional images of apple trees, and utilizing a thinning algorithm to calculate the skeleton structure. The simplified two-step pruning rule was used to identify the branches required pruning. The approach achieved pruning rates of 85% for long branches and 69% for overlapping branches of apple trees. Elfiky et al. [10] successfully acquired a comprehensive three-dimensional

Manuscript received August 15, 2024; revised January 9, 2025.

This work was supported by the Jiangxi Forestry Bureau Camellia Research Project (No. YCYJZX〔2023〕211).

Pinglu Chen is an associate professor of the College of Engineering, Jiangxi Agricultural University and Jiangxi Provincial Key Laboratory of Modern Agricultural Equipment, Nanchang, China. (e-mail: pingch757@jxau.edu.cn).

Feng Huang is a postgraduate student of the College of Software, Jiangxi Agricultural University, Nanchang, China. (e-mail: 1498508927@qq.com).

Jing Xu is a professor of the College of Engineering, Jiangxi Agricultural University and Jiangxi Provincial Key Laboratory of Modern Agricultural Equipment, Nanchang, China. (corresponding author: +8613979166680; e-mail: xujing0085@jxau.edu.cn).

Muhua Liu is a professor of the College of Engineering, Jiangxi Agricultural University and Jiangxi Provincial Key Laboratory of Modern Agricultural Equipment, Nanchang, China. (e-mail: suikelmh@sina.com).

point cloud of an apple tree by employing geometric features to register the two captured point cloud images from the front and back perspectives of the tree. They further utilized the RANSAC algorithm to accurately fit the trunk, followed by branch identification through a clustering algorithm. Notably, this method achieved a commendable branch recognition accuracy of 92.2%. Ma et al. [11] proposed a visual system based on the synchronization of two consumer RGB-D cameras to obtain high-quality three-dimensional point clouds of jujube trees on the spot. Based on the deep learning algorithm (SPGNet), the trunk and branches were automatically segmented, and then the DBSCAN clustering algorithm was used to estimate the number of branches of jujube trees. The classification accuracy of the algorithm for trunk and branch was 93% and 84% respectively. Ma et al. [12] obtained the point cloud information of jujube trees under different weather conditions through the visual system built by RGB-D camera, and then used the semantic segmentation model of Deeplabv3 and Pspnet2 to extract the mask of pruning. Finally, the pruning skeleton was extracted according to the binary image and the number of pruning branches was determined. The mPA and mIoU were 89% and 81.85%, respectively.

In summary, the identification method of fruit tree branches is mainly divided into two-dimensional image recognition and three-dimensional point cloud recognition. The accurate identification of deciduous fruit trees with sparse branches such as apples and jujubes, can be achieved through the three-dimensional reconstruction of the three-dimensional point cloud data. However, the application of these methods for identifying branches of Camellia oleifera with lush branches and evergreen characteristics poses challenges. Due to the substantial occlusion and intricate shape characteristics of Camellia oleifera tree branches, existing branch recognition methods fail to meet the precise identification and segmentation requirements for pruning operations. Therefore, this study will propose an improved YOLOv8-seg model for recognizing and segmenting Camellia oleifera branches targeted for pruning. Firstly, the SPPF module in the original backbone network is replaced by the SPPFCSPC module. This replacement strengthens the network's ability to recognize large and small branches and extract branch shape features. Secondly, the second version of deformable convolution (DCNv2) is introduced and fused with the C2f module. Additionally, some C2f modules in the Neck network are replaced, enhancing the network's capacity to capture detailed branch features, even under occlusion. Finally, the model prediction box loss function CIoU is replaced by Wise-IoU v3. This change stabilizes prediction box regression and improves the model's positioning performance.

II. Materials And Methods

A. *Data Collection and Annotation*

The present study identifies and segments the four most prevalent types of branches required pruning of Camellia oleifera, namely dead, damaged, weeping, and epicormic branches. No public dataset currently exists for the pruning requirements of Camellia oleifera branches. Therefore, this study's dataset was created by capturing images at the Camellia oleifera base in Xinjian District, Nanchang City, Jiangxi Province (LAT 116.22N, LON 28.40E), and the base in Poyang County, Shangrao City, Jiangxi Province (LAT 116.93N, LON 29.18E) using a Sony camera. The data set for training the model is divided into two categories: cloudy and sunny days, aiming to enhance its comprehensiveness. The types of branches in the data set were categorized as follows: dead branches, damaged branches, weeping branches and epicormic branches, and the number of initial data sets was 1087. The labelme [13] image annotation tool was utilized to label the data set in this study. Specifically, the polygon annotation tool within labelme was employed to accurately delineate the branch contour in the data to avoid the influence of its complex background on the branches. The tags of dead branches, damaged branches, weeping branches and epicormic branches were set as dead branch, damaged branch, weeping branch and epicormic branch respectively. The Json tag file generated after labeling was converted into txt file.

B. *Data Augmentation*

Data augmentation is crucial in deep learning. It expands datasets, addresses class imbalance, reduces overfitting risk, and enhances the model's generalization ability. Due to the small number of data sets, this study expanded the data set by data enhancement methods. The data enhancement methods include brightness change, increase noise, add random points and mirror reflection, as shown in Fig 1. The expanded data set comprises a total of 5435 images, with the training sets, verification sets, and test sets distributed in an 8:1:1 ratio. The distribution of branches by type is presented in Table I.

Fig. 1. Data augmentation

TABLE I
CAMELLIA OLEIFERA BRANCHES TO BE PRUNED AND ITS NUMBER

	Number of dead branches	Number of damaged branches	Number of weeping branches	Number of epicormic branches
Sunny day dataset (initial)	142	140	149	105
Cloudy day dataset (initial)	145	150	148	108
Sunny day dataset (expansion)	710	700	745	525
Cloudy day dataset (expansion)	725	750	740	540
Training set	1135	1150	1185	877
Validation set	150	150	150	94
Testing set	150	150	150	94
Grand total	1435	1450	1485	1065

C. *Introduction of YOLOv8-seg Network Model*

The Ultralytics team has developed the YOLOv8 algorithm in recent years as part of the ongoing YOLO series [15]-[16]-[17]-[18]. This algorithm demonstrates exceptional capabilities in various computer vision tasks, including object detection, image classification, and instance segmentation. In the backbone and Neck networks, YOLOv8 adopts the design concept of YOLOv7 ELAM [19]-[20]. It replaces the C3 structure in YOLOv5 [21]-[22] with the C2f structure, which provides richer gradient flow, and adjusts channel numbers for different scale models. In the head, the original coupling head is replaced with a decoupling head, and the anchor-based approach is switched to anchor-free, significantly enhancing model performance. YOLOv8-seg is an instance segmentation model in YOLOv8, which is specially designed for instance segmentation tasks. The segmentation principle is based on the principle of YOLACT network. YOLOv8-seg has five types of network structures, namely YOLOv8n-seg, YOLOv8s-seg, YOLOv8m-seg, YOLOv8l-seg and YOLOv8x-seg. In this study, YOLOv8n-seg with the smallest model structure was selected. The network structure of YOLACT is shown in Fig 2.

D. *Optimization of YOLOv8n-seg Network Model*

1) *Introduction of SPPFCSPC Module*

The original YOLOv8 model struggles with detection and segmentation due to the occlusion and similar shape characteristics of Camellia oleifera branches. To address this, the SPPF module in the backbone network's tail was replaced with the SPPFCSPC module. The SPPFCSPC module is improved on the basis of the SPPCSPC module used in YOLOv7. The SPPFCSPC module performs four distinct Max Pool operations with varying kernel sizes in its first branch. This design enhances its ability to handle diverse objects and improves discrimination between large and small targets within the image. This improved performance surpasses that of the SPPF module, albeit at a relatively increased parameter count. The SPPFCSPC module is a novel integration of the SPPCSPC and SPPF modules, harnessing their respective strengths to enhance computational speed while preserving the same receptive field. The network architectures of SPPCSPC and SPPFCSPC are illustrated in Fig 3 and Fig 4, correspondingly.

2) *C2f Network Combined with Deformable Convolution*

The task of detecting and segmenting branches for pruning in Camellia oleifera trees presents unique and intricate features. The diverse shape of the branches to be pruned coupled with the obstructive presence of leaves, pose significant challenges for effectively extracting features by the model. In the original model's feature extraction stage, traditional convolution may fail to capture intricate branch features, potentially causing false detections. To address this limitation, this study introduced the second edition of deformable convolution (DCNv2) [23] to eliminate the shortcomings of traditional convolution in the feature extraction stage. The second version of deformable convolution is based on deformable convolution (DCN). While deformable convolution (DCN) adapts to geometric changes by adjusting sampling positions based on input features, some sampling points extend beyond the region of interest, impairing feature extraction.

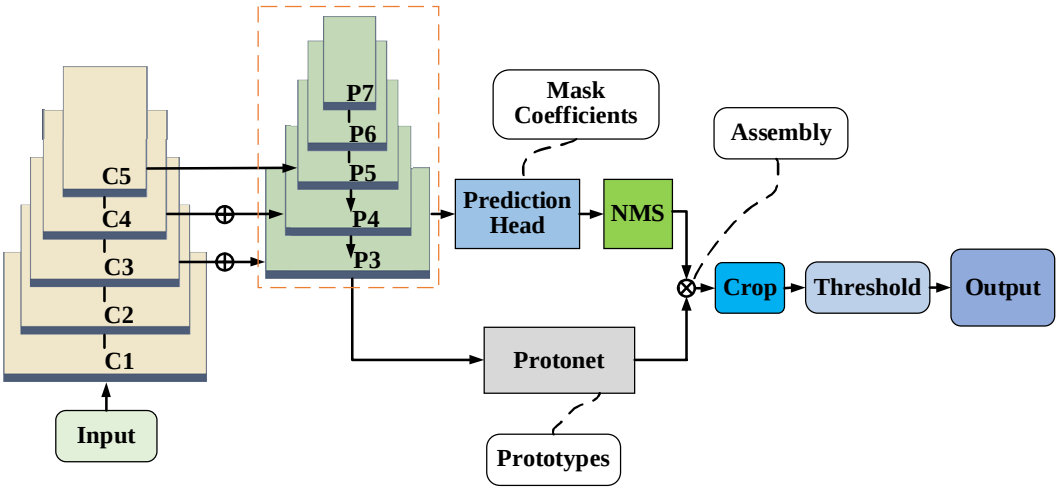

Fig. 2. YOLACT network structure flow chart

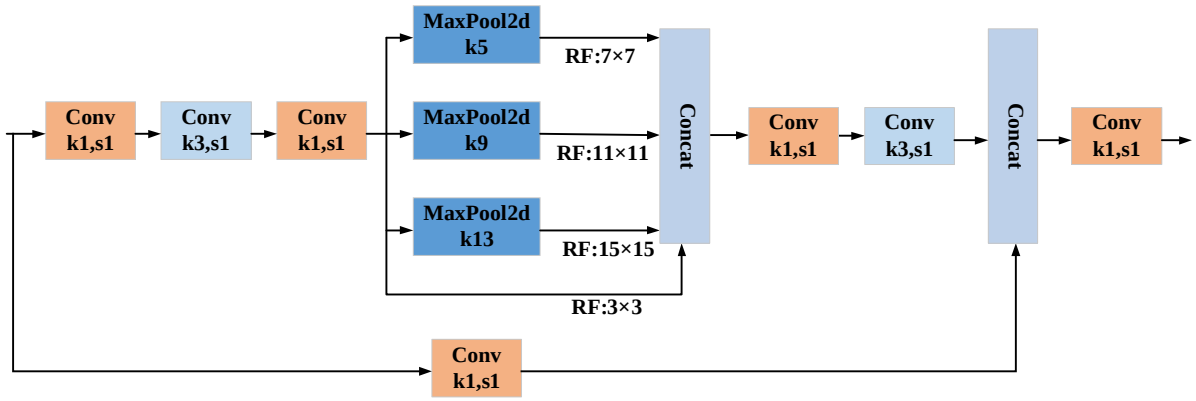

Fig. 3. SPPCSPC network structure

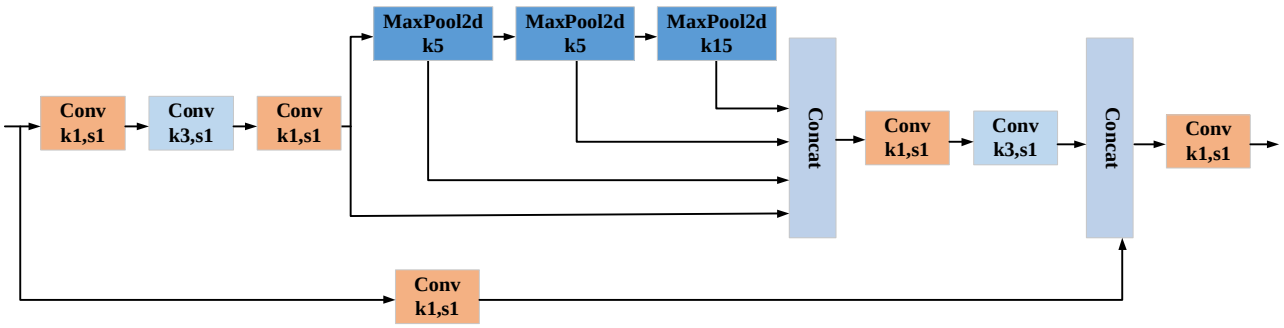

Fig. 4. SPPFCSPC network structure

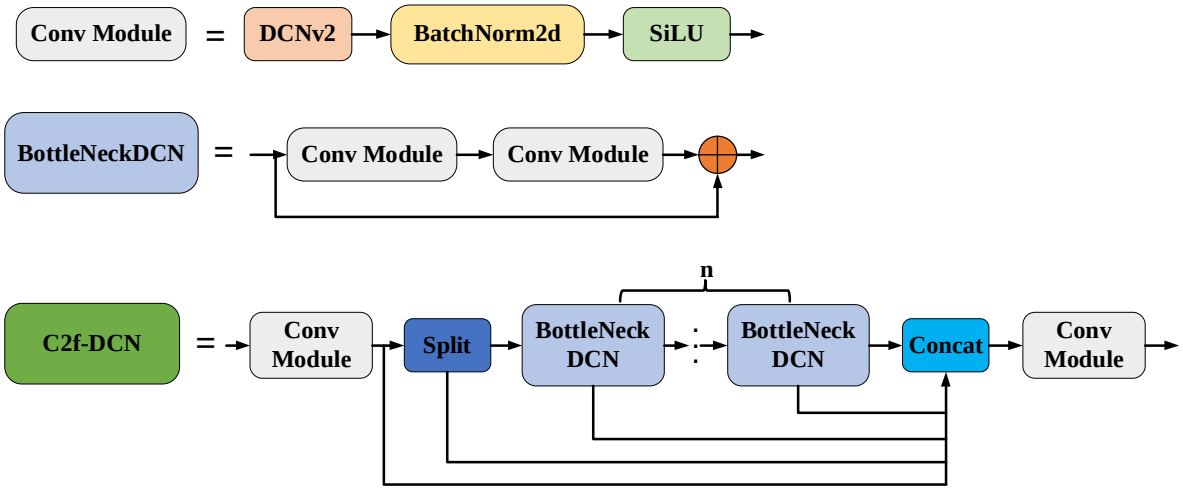

Fig. 5. C2f-DCN module network structure

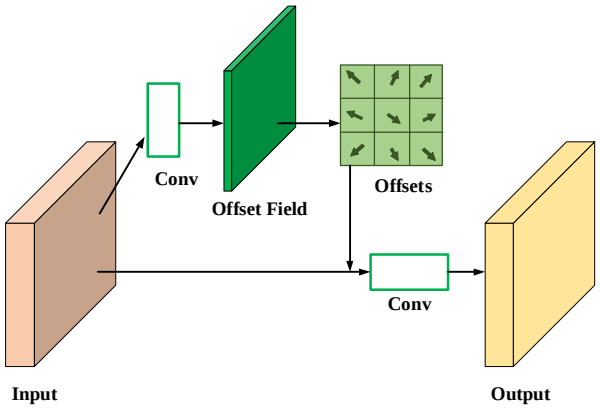

Fig. 6. DCNv2 network structure

The second edition of deformable convolution introduces a modulation mechanism that can stack more deformable convolution layers, and proposes a feature simulation scheme to guide network training, thus eliminating the shortcomings of deformable convolution. However, only adding the second version of the deformable convolution does not significantly improve the performance of the model. This study designed the C2f-DCN module based on the C2f network and integrated it into the 15th layer of the Neck network. Experimental results reveal that excessive replacement of C2f-DCN modules increases model parameters without significantly improving performance. However, carefully replacing specific C2f-DCN modules effectively addresses the occlusion issue in identifying and segmenting Camellia oleifera branches for pruning. The C2f-DCN module network structure and the DCNv2 network structure are shown in Fig 5 and Fig 6, respectively.

3) *Optimization of Loss Function*

YOLOv8 uses CIoU Loss as its prediction box loss function. CIoU loss is an improvement over DIoU Loss, and it incorporates factors such as the distance between the centers of the prediction and ground truth boxes and their aspect ratio. These additions enhance the stability of box regression. However, the method for calculating the aspect ratio between prediction and ground truth boxes is not clearly defined, which negatively impacts network optimization. To address this, this study proposes replacing CIoU Loss with Wise-IoU v3 [24]. Wise-IoU v3 is an improved version of Wise-IoU v1. It introduces distance attention based on distance metrics, thereby forming a dual-attention mechanism that enhances the model's ability to focus on relevant features. The specific formula is as follows:

$$L_{\mathrm{WIoUv1}}=R_{\mathrm{WIoU}}L_{\mathrm{IoU}} \tag{1}$$

$$R_{\mathrm{WIoU}}=\exp(((x-x_{\mathrm{gt}})^2+(y-y_{\mathrm{gt}})^2)/(W_{\mathrm{g}}^2+H_{\mathrm{g}}^2)^{*}) \tag{2}$$

Note: L_{WIoUv1} is the bounding box loss of Wise-IoU v1; L_{IoU} is the bounding box loss of IoU; R_{WIoU} is the distance attention; x and y are the abscissa and ordinate of the center point of the prediction box, respectively; x_{gt} and y_{gt} are the abscissa and ordinate of the center point of the real box, respectively; W_{g} and H_{g} are the width and height of the minimum circumscribed rectangle of the prediction box and the real box, respectively; * represents the separation operation from the calculated graph to make it a constant without gradient.

The Wise-IoU v3 loss function introduces the concept of outlier degree to evaluate the quality of anchor boxes. Based on differences in outlier degree, corresponding gradient enhancements are assigned. This approach constructs a non-monotonic focus coefficient, which is integrated into the Wise-IoU v1 loss function. As a result, Wise-IoU v3 can dynamically adjust gradient allocation in different scenarios, improving the model's localization accuracy. The specific formula for Wise-IoU v3 is as follows:

$$\beta=L_{\mathrm{IoU}}^{*}/\overline{L_{\mathrm{IoU}}} \tag{3}$$

$$L_{\mathrm{WIoUv3}}=rL_{\mathrm{WIoUv1}} \tag{4}$$

$$r=\beta/\delta\alpha^{\beta-\delta} \tag{5}$$

Note: L_{IoU}^{*} is a monotone focusing coefficient; β is an outlier and $\beta \in [0, +\infty]$; $\overline{L_{\mathrm{IoU}}}$ is a moving average; r is a nonmonotone focusing factor; α and δ are hyperparameters, α= 1.9, δ= 3.

4) *Improved YOLOv8n-seg Network Model*

Building on the previous analysis, this study enhanced the YOLOv8n-seg model through several key modifications. First, the SPPF module in the original backbone network was replaced with the SPPFCSPC module, thereby improving the network's ability to recognize both large and small branches in images and extract branch shape features more effectively. Second, the second version of deformable convolution (DCNv2) was integrated into the C2f module, replacing several C2f modules in the Neck network. This change enables the model to better capture detailed features of occluded branches, thus enhancing overall performance. Finally, the original CIoU loss function for the prediction box was substituted with Wise-IoU v3, leading to more stable box regression and improved positioning accuracy. The enhanced YOLOv8n-seg model is illustrated in Fig 7.

E. *Evaluation Index*

In order to objectively analyze the classification and segmentation performance of the Camellia oleifera tree branch data set, this study introduced evaluation indicators such as precision (P), recall (R), mean average precision (mAP), F1-score and inferring time per image. The primary objective of this study is to accurately identify and segment the branches that require pruning. Therefore, the mean average precision (mAP) was employed as the main evaluation index with the IoU threshold set at 0.5. The evaluation indicators can be expressed as follows:

$$P=TP/(TP+FP)\times 100\% \tag{6}$$

$$R=TP/(TP+FN)\times 100\% \tag{7}$$

$$F1\text{-}score=2\cdot P\cdot R/(P+R)\times 100\% \tag{8}$$

$$mAP=\frac{1}{\mathrm{C}}\sum_{k=i}^{\mathrm{N}}\mathrm{P}(k)\,\Delta R(k)\times 100\% \tag{9}$$

Note: *TP* refers to the accurate identification of positive samples as positive samples; *FP* indicates the number of negative samples incorrectly identified as positive samples. *FN* indicates the number of positive samples incorrectly identified as negative samples; *C* is the number of categories; *N* is the number of reference thresholds; *k* is the threshold; $P(k)$ is the accuracy rate; $R(k)$ is the recall rate.

III. Model Training And Result Analysis

A. *Experimental Environment and Parameter Settings*

The training environment was set up on a Windows 10 operating system with an NVIDIA GeForce RTX 3080 GPU. The system included 32GB of RAM, 10GB of GPU memory, and an Intel Core i7-12700F CPU. CUDA version 11.3 and Python 3.8.16 were used as development tools. The model training involved 200 iterations with a batch size of 32 per iteration. The initial learning rate was set to 0.01, the momentum parameter was set to 0.937, and the weight decay factor was set to 0.0005.

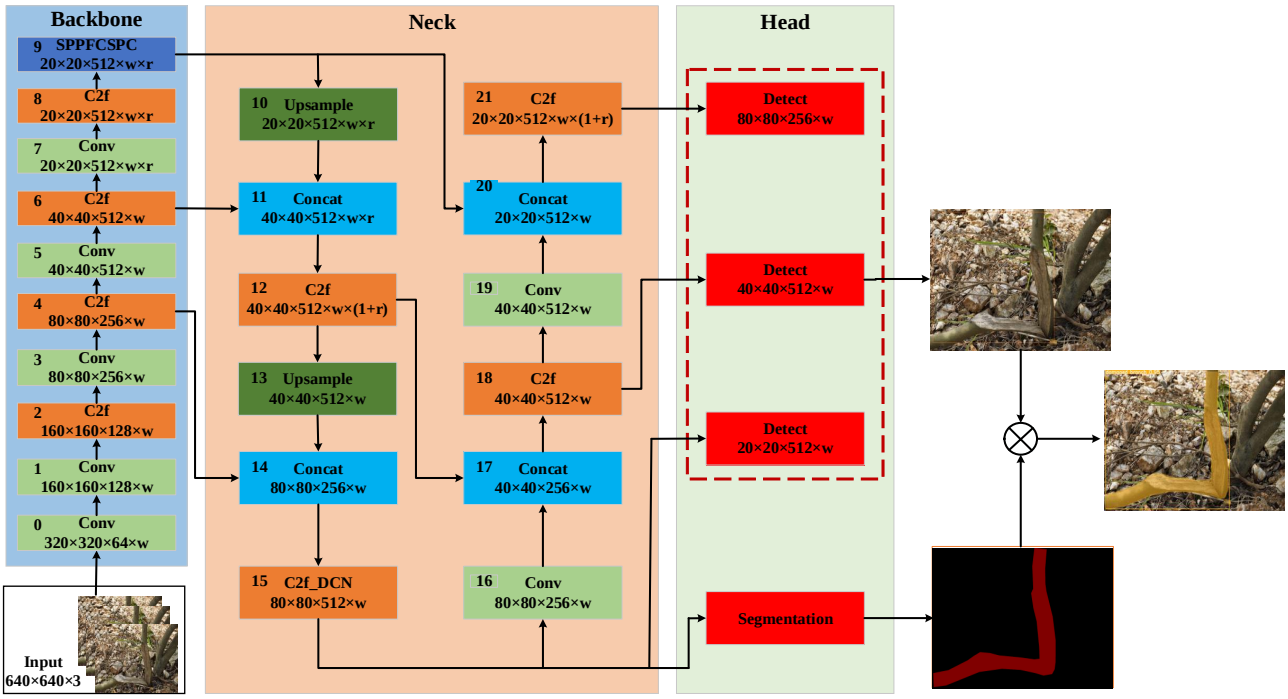

Fig. 7. Improve the network structure of YOLOv8n-seg

B. *Analysis of Model Results*

The improved YOLOv8n-seg model was assessed for its effectiveness in accurately identifying and segmenting the branches to be pruned of Camellia oleifera. A total of 4347 images from the test set were utilized for training, and a comparison was made against the original model. The results obtained from the testing phase are presented in Fig 8 and Fig 9.

As illustrated in Fig 8, the loss value decreases most rapidly during the first 25 epochs for both the original and improved models. Subsequently, it gradually converges to a stable state after 100 epochs. The model is configured to terminate the mosaic data at the final 10 epochs during training, resulting in a subsequent decrease in loss value and eventual stabilization. The box-loss and mask-loss values of the improved model are significantly lower than those of the original model, indicating that the model improvement is effective. Fig 9 reveals that the mAP value experiences the most rapid increase during the initial 25 epochs following model improvement, gradually stabilizing after reaching 100 epochs. Notably, the improved model exhibits significantly higher box-mAP and mask-mAP compared to its original counterpart, reaffirming the effectiveness of our proposed enhancements.

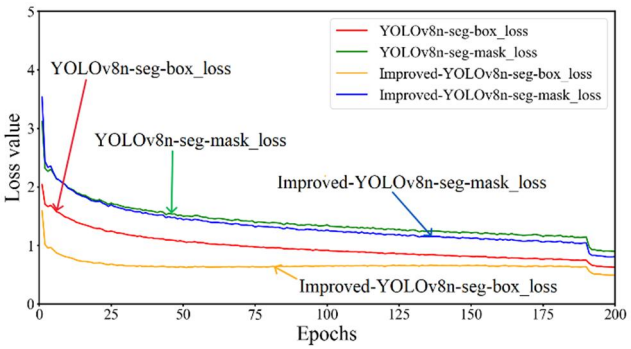

Fig. 8. Loss function curve before and after the model improvement

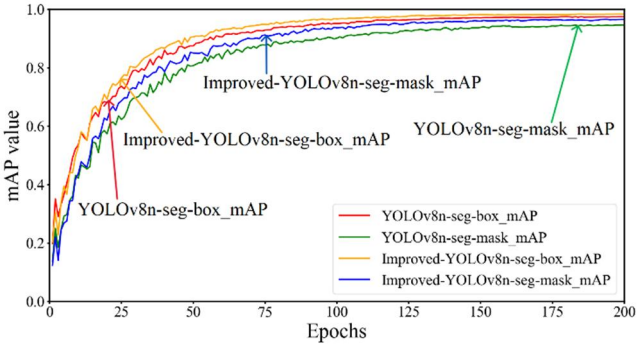

Fig. 9. mAP curve before and after the model improvement

C. *Comparative Experiments on Different Pyramid Pooling Structures*

The original YOLOv8n-seg model employs the SPPF module for pyramid pooling. To evaluate the efficacy and versatility of the SPPFCSPC module introduced in this study, a comparative analysis was conducted with several commonly used pyramid pooling modules, including the SimSPPF, ASPP, and SPPCSPC modules. Each of these five modules, along with the original SPPF module, was integrated into the YOLOv8n-seg model, and the data set of Camellia oleifera to be pruned was utilized for training and validation. The evaluation results of the experiment are shown in Table II.

TABLE II
COMPARISON OF EVALUATION INDEXES OF DIFFERENT PYRAMID POOLING STRUCTURES

Network	mAP/%		F1-score/%		Model size/MB
	box	mask	box	mask	
SPPF	96.8	94.8	94.53	93.07	6.45
SimSPPF	96.6	94.1	93.9	92.42	6.45
ASPP	96.8	93.8	94.49	92.64	10.3
SPPCSPC	97.4	95.3	95.21	93.65	9.53
SPPFCSPC	97.4	95.6	95.23	93.82	9.53

As illustrated in Table II, integrating both the SPPCSPC and SPPFCSPC modules leads to improvements in mAP and F1-score compared to the original model with only the SPPF module. However, this enhancement is accompanied by an increase in model size of 3.08 MB. In contrast, the SPPFCSPC module enhances mAP and F1-score without increasing the model size. Specifically, the introduction of the SPPFCSPC module results in a 0.6% increase in box-mAP, a 0.8% increase in mask-mAP, a 0.7% increase in box F1-score, and a 0.75% increase in mask F1-score relative to the original model. These findings underscore the effectiveness of incorporating the SPPFCSPC module for improving model performance.

D. *Improved Model Ablation Experiment*

The SPPFCSPC module was introduced in this study based on YOLOv8n-seg, enhancing the network's recognition ability for both large and small branches as well as its extraction capability of branch shape features during the feature extraction stage. Additionally, the C2f module in the Neck network incorporated the second version of deformable convolution (DCNv2) to replace some existing C2f modules, thereby improving the network's ability to capture detailed features of occluded branches that need to be pruned. Finally, Wise-IoU v3 replaced the prediction box regression loss function, ensuring a more stable regression process. To assess the impact of the three improvement strategies on model performance, ablation experiments were conducted. The results are presented in Table III.

The original YOLOv8n-seg model, as indicated in Table III, was employed for experiment 1. It achieved the following results: a box-mAP of 96.8%, a mask-mAP of 94.8%, a box F1-score of 94.53%, a mask F1-score of 93.07%, and a model size of 6.45 MB. Experiment 2 introduced the SPPFCSPC module on the basis of experiment 1 to enhance the ability of the model to identify large and small branches and extract the shape features of branches. In Experiment 2, the box-mAP, mask-mAP, box F1-score, and mask F1-score increased by 0.6%, 0.8%, 0.7%, and 0.75%, respectively, compared to Experiment 1. However, this improvement came at the cost of a 3.08 MB increase in model size due to additional parameters. In Experiment 3, DCNv2 was integrated into the C2f module based on the architecture from Experiment 2, enhancing the model's capability to capture detailed branch features under occlusion. The box-mAP, mask-mAP, box F1-score and mask F1-score of the model increases by 0.1%, 0.4%, 0.15% and 0.41%, respectively, and the model size increases by 0.03 MB. In experiment 4, based on experiment 3, the loss function of calculating the prediction box regression was replaced by Wise-IoU v3. The box-mAP, box F1-score and mask F1-score of the model are increased by 0.4%, 0.66% and 0.4%, respectively. Because the loss function of calculating the prediction box regression is replaced, the mask-mAP of the model does not increase, and the relative model size does not change. Through the above comparative analysis, this study has positive significance for all improvements of YOLOv8n-seg model.

To evaluate the performance improvements of the enhanced model relative to the original, four images were randomly selected from the test set for detailed analysis (Fig 10). As illustrated in Fig 10, the enhanced model effectively addresses issues related to incomplete segmentation of epicormic branch edges and repetitive recognition during segmentation. Additionally, it improves the accuracy of identifying dead branches by reducing both missed detections and redundant segmentations. These enhancements underscore the practical value and reliability of the improved model.

E. *Comparative Experiments of Different Models*

The improved model was compared to several mainstream instance segmentation models, including YOLOv5n-seg, Mask R-CNN [25], Mask Scoring R-CNN (MS R-CNN) [26], Hybrid Task Cascade (HTC) [27] and YOLOv9-seg. Convergence training was conducted for these models and their performance was compared as presented in Table IV.

The improved YOLOv8n-seg model achieves a box-mAP of 97.9% according to Table IV, surpassing the performance of YOLOv5n-seg model, Mask R-CNN model, and MS R-CNN model by 2.3%, 3.2%, and 3.3% respectively. It only exhibits a slight decrease in performance compared to the HTC model (0.1%) and the YOLOv9-seg model (0.5%). The improved model achieves a mask-mAP that is respectively 12.9%, 6%, 6.3%, and 0.7% higher than YOLOv5n-seg, Mask R-CNN, MS R-CNN, and HTC models while being only 0.7% lower than the YOLOv9-seg model. In terms of inference time, the improved YOLOv8n-seg model exhibits an inference time of 39.5ms, which is only 17.1ms slower than the YOLOv5n-seg model but still maintains significant advantages over other instance segmentation models. The improved model has a compact size of only 9.56 MB, which provides distinct advantages over other instance segmentation models and facilitates efficient deployment in practical applications. In summary, the improved YOLOv8n-seg model exhibits superior comprehensive advantages in terms of accuracy, inference time, and model size, thereby better fulfilling the requirements for rapid and precise identification and segmentation of Camellia oleifera branches to be pruned within the intricate environment of a Camellia oleifera forest.

TABLE III
IMPROVED MODEL ABLATION EXPERIMENTAL RESULTS

Order	SPPFCSPC module	C2f-DCN module	Wise-IoU v3 loss	mAP/%		F1-score/%		Model size / MB
				box	mask	box	mask	
1	-	-	-	96.8	94.8	94.53	93.07	6.45
2	ü	-	-	97.4	95.6	95.23	93.82	9.53
3	ü	ü	-	97.5	96	95.38	94.23	9.56
4	ü	ü	ü	97.9	96	96.04	94.63	9.56

Note: The black ellipse region represents incomplete segmentation, the green ellipse region represents missed detection, and the yellow ellipse region represents repeated segmentation.

Fig. 10. Comparison before and after model improvement

TABLE IV
COMPARISON RESULTS OF DIFFERENT MODELS

Model	mAP/%		Inference time/ms	Model size/MB
	box	mask		
YOLOv5n-seg	95.6	83.1	22.4	3.92
Mask R-CNN	94.7	90	86.21	335
MS R-CNN	94.6	89.7	93.28	459
HTC	98	95.3	149.25	588
YOLOv9-seg	98.4	96.7	66.7	111
This study	97.9	96	39.5	9.56

IV. DISCUSSION

The accurate identification and segmentation of branches to be pruned are crucial in pruning operations for fruit and Camellia oleifera trees because they enable the precise localization of the pruning point for the pruning robot. The working environment in orchards and Camellia oleifera forests is highly complex. Branches exhibit diverse shapes and often overlap, which significantly affects recognition accuracy. Qiao et al. [28] proposed an enhanced PSPNet network model for precise identification and segmentation of jujube tree trunks, achieving an impressive IoU score of 81.88%. This approach effectively satisfies the requirements for subsequent tasks in jujube harvesting robotics. Similarly, Zheng et al. [8] introduced a refined Ghost-HRNet network model to precisely recognize and segment jujube tree trunks and branches, achieving a remarkable mPA score of 89.46%. Compared to other semantic segmentation models, it exhibits distinct advantages, thereby providing valuable theoretical support for the advancement of future jujube harvesting robots. In contrast to the aforementioned research objects, Yang et al. [29] proposed a citrus branch recognition and segmentation method based on the Mask RCNN model for reconstructing citrus branches. This approach achieved an average reconstruction accuracy of 88.64% for fruit tree branches, thereby providing theoretical support for subsequent studies on obstacle avoidance picking by robotic systems.

The above analysis indicates that branch recognition and segmentation in orchards are frequently examined using semantic or instance segmentation models. The primary challenge of this study lies in accurately identifying and segmenting four distinct types of pruning branches among all Camellia oleifera branches. Unlike previous research that primarily focused on single targets such as individual branches or stems, this study comprehensively analyzes and segments four common types of pruning branches found in Camellia oleifera trees. These branches exhibit growth patterns similar to those of normal branches, which poses significant challenges for accurate machine identification and segmentation.

To address these challenges, an enhanced YOLOv8n-seg model was proposed based on a comprehensive analysis of previous research. This model not only simultaneously identifies and segments various pruning branches but also surpasses the Mask RCNN model in terms of accuracy, inference time, and model size. It is well-suited for complex orchard environments and provides a valuable reference for future studies on intelligent machinery operations in such settings.

V. CONCLUSION

This study aims to develop a method for accurately and efficiently identifying and segmenting pruning branches in Camellia oleifera trees. It establishes a theoretical basis for precise pruning point positioning by pruning robots. The

results demonstrate the superior accuracy and efficiency of a deep learning-based instance segmentation model in identifying and segmenting pruning branches. The specific conclusions are summarized as follows:

1) To address the challenges of similar shapes and mutual occlusion between pruning and normal branches of Camellia oleifera in natural environments, this study introduced the SPPFCSPC module, C2f-DCN model, and Wise-IoU v3 loss function. The SPPFCSPC module improved the model's ability to identify branches of varying sizes and extract shape features. As a result, the model achieved a box-mAP of 97.4% and a mask-mAP of 95.6%, surpassing the original model by 0.6% and 0.8%, respectively. Additionally, the inclusion of the C2f-DCN model and Wise-IoU v3 loss function enhanced the model's capability to capture intricate features of occluded branches. This led to further improvements of 1.1% in box-mAP and 1.2% in mask-mAP.

2) The improved YOLOv8n-seg model, along with other leading instance segmentation models such as YOLOv5n-seg, YOLOv9-seg, Mask R-CNN, MS R-CNN, and HTC, was trained and evaluated on a self-created dataset for camellia tree pruning. This process validated the superiority of the improved model. The experimental results demonstrate that the improved YOLOv8n-seg model achieves a box-mAP of 97.9% and a mask-mAP of 96% on the test set. The inference time for processing a single image is measured to be 39.5 ms, while the model size amounts to 9.56 MB. The improved model exhibits higher accuracy compared to the original model. Compared to other mainstream instance segmentation models, the improved YOLOv8n-seg exhibits clear advantages in comprehensive indices. It effectively addresses recognition and segmentation tasks for Camellia oleifera pruning within real forest environments. It presents a viable solution for the subsequent investigation of mechanized pruning equipment for Camellia oleifera trees.

However, this method still possesses certain limitations and drawbacks. Firstly, the validation of this approach is solely conducted through computer simulations before its practical implementation on operational machinery. Secondly, in the process of pruning Camellia oleifera trees, there exists a wide range of branch types that extend beyond the four categories examined in this study. Therefore, further investigation is required to address the identification and segmentation challenges associated with other branches targeted for pruning.

REFERENCES

[1] L. Zhang and L. Wang, “Development status and prospect of Camellia oleifera industry in China,” China Oils Fats, pp. 6-9, 2021.

[2] Y. Chen, “On the pruning and shaping of Camellia oleifera,” For. Ecol, pp. 34-36, 2018.

[3] J. Sun, “Reshaping and pruning of high-yield Camellia oleifera,” Mod. Agric. Sci. Tech, pp. 175-176, 2018.

[4] T. Jia, Y. Xun, G. Bao, M. Dong, and Q. Yang, “Research on grape branch skeleton extraction algorithm based on machine vision,” J. Mech. Electr. Eng, pp. 501-504, 2013.

[5] B. Huang, M. Shao, and L. Song, “Visual recognition and frame extraction of loquat branch pruning robot,” J. South China Univ. Technol., Nat. Sci, pp. 114-119, 2015.

[6] S. Amatya, M. Karkee, A. Gongal, Q. Zhang and M. D. Whiting, “Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting,” Biosyst. Eng. 146, pp. 3-15, 2016.

[7] J. Zhang, L. He, M. Karkee, Q. Zhang, X. Zhang and Z. Gao, “Branch detection for Apple Trees Trained in Fruiting Wall Architecture Using Depth Features and Regions-Convolutional Neural Network (R-Cnn),” Comput. Electron. Agr,155, pp. 386-393, 2018.

[8] Z. Zheng, Y. Hu, T. Guo, Y. Qiao, Y. He, Y. Zhang and Y. Huang, “AGHRNet: An attention ghost-HRNet for confirmation of catch-and-shake locations in jujube fruits vibration harvesting,” Comput. Electron. Agr, 210, p. 107921, 2023.

[9] M. Karkee, B. Adhikari, S. Amatya and Q. Zhang, “Identification of pruning branches in tall spindle apple trees for automated pruning,” Comput. Electron. Agr, 103, pp. 127-135, 2014.

[10] N. M. Elfiky, S. A. Akbar, J. Sun, J. Park and A. Kak. “Automation of dormant pruning in specialty crop production: An adaptive framework for automatic reconstruction and modeling of apple trees”, IEEE, pp. 65-73, 2015.

[11] B. Ma, J. Du, L. Wang, H. Jiang and M. Zhou, “Automatic branch detection of jujube trees based on 3D reconstruction for dormant pruning using the deep learning-based method”, Comput. Electron. Agr, 190, p. 106484, 2021.

[12] B. Ma, J. Yan, L. Wang, and H. Jiang, “Pruning recognition and skeleton extraction of dwarf and dense jujube trees based on semantic segmentation”, Trans. Chin. Soc. Agric. Mach, 53, pp. 313-319, 2022.

[13] B. C. Russell, A. Torralba, K. P. Murphy and W. T. Freeman, “LabelMe: A Database and Web-Based Tool for Image Annotation”, Int. J. Comput. Vision 77, pp. 157-173, 2008.

[14] X. Yue, K. Qi, X. Na, Y. Zhang, Y. Liu and C. Liu, “Improved yolov8-seg network for instance segmentation of healthy and diseased tomato plants in the growth stage”, Agriculture 13(8), p. 1643, 2023.

[15] A. Bochkovskiy, W. Chien-Yao and M. L. Hong-Yuan, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, Cornell University Library, arXiv.org, Ithaca, 2020.

[16] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement”, Cornell University Library, arXiv.org, Ithaca, 2018.

[17] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger”, Cornell University Library, arXiv.org, Ithaca, 2016.

[18] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei and X. Wei, “YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications”, Cornell University Library, arXiv.org, Ithaca, 2022.

[19] W. Chien-Yao, A. Bochkovskiy and M. L. Hong-Yuan, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”, Cornell University Library, arXiv.org, Ithaca, 2022.

[20] J. Fu, and Y. Tian, “Improved YOLOv7 Underwater Object Detection Based on Attention Mechanism”, Engineering Letters, vol. 32, no. 7, pp1377-1384, 2024.

[21] R. Shan, X. Zhang, and S. Li, “A Method of Pneumonia Detection Based on an Improved YOLOv5s,” Engineering Letters, vol. 32, no. 6, pp1243-1254, 2024.

[22] Z. Lin, L. Zhu, J. Zhang, Y. Zhang, and X. Liu, “Research on Improving YOLOv5s Algorithm for Defect Detection in Cylindrical Coated Lithium-ion Batteries,” Engineering Letters, vol. 32, no. 7, pp1521-1528, 2024.

[23] X. Zhu, H. Hu, S. Lin and J. Dai, “Deformable ConvNets v2: More Deformable, Better Results”, Cornell University Library, arXiv.org, Ithaca, 2018.

[24] Z. Tong, Y. Chen, Z. Xu and R. Yu, “Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism”, Cornell University Library, arXiv.org, Ithaca, 2023.

[25] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn”, In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988, 2017.

[26] Z. Huang, L. Huang, and Y. Gong, “Mask scoring r-cnn”, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6409–6418, 2019.

[27] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang, C. C. Loy and D. Lin, “Hybrid Task Cascade for Instance Segmentation”, Cornell University Library, arXiv.org, Ithaca, 2019.

[28] Y. Qiao, Y. Hu, Z. Zheng, Z. Qu, C. Wang, T. Guo and J. Hou, “A Diameter Measurement Method of Red Jujubes Trunk Based on Improved PSPNet”, Agriculture, vol. 12, no. 8, pp. 1140, 2022.

[29] C. Yang, Z. Wang, L. Xiong, Y. Liu, X. Kang and W. Zhao, “Identification and Reconstruction of Citrus Branches under Complex Background Based on Mask R-CNN” Trans. Chin. Soc. Agric. Mach, 50, pp. 22-30, 69, 2019.