A MELON FRUIT DIAMETER MEASUREMENT METHOD BASED ON AN IMPROVED MASK R-CNN

/

一种基于改进 MASKRCNN 的甜瓜果径测量方法

Deyang LYU, Xincheng LI*), Weidong WANG, Baorong WU, Shenghao SHI, Huiyong SHEN
College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Shandong / China
E-mail: xincheng_li@163.com
Corresponding author: XinCheng Li
DOI: https://doi.org/10.35633/inmateh-77-09

Keywords: *Binocular vision, Fruit diameter, Mask R-CNN, Melon, Non-contact measurement*

ABSTRACT

Measuring melon fruit diameter offers key insights into growth status and maturity. To overcome the limitations of manual measurement—namely high labor demands, time consumption, and large errors—this study introduces a method based on an improved Mask R-CNN algorithm. The model uses ResNet50 as the backbone and incorporates a Channel Prior Convolutional Attention (CPCA) mechanism and a bidirectional feature fusion pyramid network to enhance multi-scale feature extraction. A Self-Attention (SE) mechanism is added to the mask branch to improve segmentation accuracy. Measurement points are determined through contour segmentation, curvature analysis, and bounding rectangle fitting. A binocular camera provides depth information, and Euclidean distance is used to compute actual size. The improved algorithm achieves detection and segmentation precision of 94.2% and 92.7%, with recall rates of 94.5% and 93.6%. The method yields average relative errors of 7.1% (horizontal) and 7.6% (vertical), meeting practical agricultural needs and supporting maturity assessment.

摘要

测量甜瓜果实直径可以提供对生长状态和成熟度的关键见解。为了克服人工测量的局限性，即高劳动力需求、时间消耗和大误差，本研究引入了一种基于改进的 Mask R-CNN 算法的方法。该模型使用 ResNet50 作为骨干，并结合了信道先验卷积注意力（CPCA）机制和双向特征融合金字塔网络，以增强多尺度特征提取。在掩码分支中添加了自注意（SE）机制以提高分割精度。通过轮廓分割、曲率分析和边界矩形拟合来确定测量点。双目相机提供深度信息，欧几里德距离用于计算实际尺寸。改进后的算法实现了 94.2% 和 92.7% 的检测和分割精度，召回率分别为 94.5% 和 93.6%。该方法的平均相对误差为 7.1%（水平）和 7.6%（垂直），满足实际农业需求，支持成熟度评估。

INTRODUCTION

The primary parameters of melon fruit diameter include vertical and horizontal diameters *(Chang et al., 2018)*, which reflect the growth status of the fruit and are important indicators of fruit maturity *(Xue et al., 2024; Gothi et al., 2022)*. Currently, melon diameter measurement mainly relies on manual methods, which are inefficient, labor-intensive, and prone to subjective errors *(Wang et al., 2024)*. Existing machine vision-based approaches often depend on traditional image preprocessing and edge detection techniques, which are sensitive to lighting, background conditions, and fruit phenotype, limiting their practical application and leading to significant errors.

Mask R-CNN, a deep learning model evolved from Faster R-CNN, enables both object detection and pixel-level segmentation, and has been widely used in various fields. *He et al., (2018), Gu et al., (2024) and Geng et al., (2022),* improved the RT-DETR model to enhance tomato detection precision and used edge detection combined with Hough transform to measure tomato diameters, aiding in fruit detection and localization for automatic picking robots. *Basak et al., (2022),* developed a simple linear regression model using pixel count to estimate strawberry weight. *Zheng et al., (2021)* proposed an automated 3D point cloud registration algorithm. By utilizing a Kinect camera and a parameter optimization method, they achieved high-precision and efficient reconstruction of butterhead lettuce, with an average error of 6.5 mm, demonstrating the feasibility of using binocular vision technology for plant phenotype measurement. Despite advances in deep learning and machine vision, studies on melon diameter measurement remain limited.

This research focuses on melon diameter measurement. The proposed method enhances Mask R-CNN for better detection and segmentation of melon fruit, identifies pixel coordinates of diameter measurement points from the segmented contour, retrieves depth information using binocular vision to convert pixel to 3D coordinates, and finally computes fruit diameter through Euclidean distance, enabling non-contact measurement and providing data for fruit maturity assessment.

MATERIALS AND METHODS

Dataset Construction and Annotation

The melon images used in this study were collected from a greenhouse located within the Shanhou Renjia Agricultural Complex in Laixi, Qingdao, Shandong Province, China (120°39′ E, 36°72′ N). The melons were cultivated using a vertical hanging vine method, and the fruit maturity ranged from the fruit enlargement stage to full ripeness. Image acquisition was conducted on June 1, 2024, using a 64-megapixel digital camera. The imaging distance ranged from 30 to 100 cm, and all images were saved in JPG format. A total of 500 images were collected and annotated using the Labelme software, with annotations saved in JSON format. To enhance the robustness of model training, the original images and annotations were augmented using methods such as noise addition, brightness adjustment, spatial shifting, and rotation. This resulted in a final dataset of 2,000 images, which was split into a training set of 1,600 images and a test set of 400 images in an 8:2 ratio. The dataset construction process is illustrated in Figure 1.

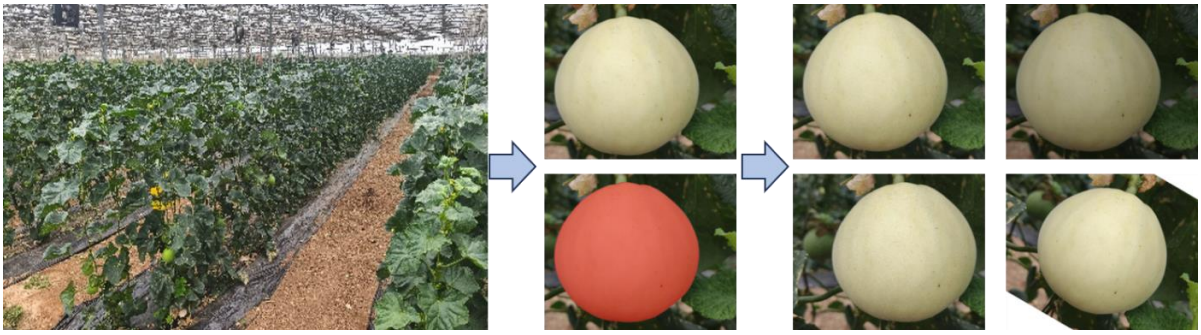


Fig. 1 - Dataset Construction Process

Improved Mask R-CNN Algorithm

The experiments were conducted on a portable computer equipped with an Intel(R) Core (TM) i5-10300H CPU @ 2.50 GHz, 16 GB of RAM, 465 GB of storage, and an NVIDIA GeForce GTX 1660 Ti GPU. The improved Mask R-CNN model was developed, trained, and evaluated using the Detectron2 framework based on PyTorch.

Mask R-CNN extends object detection into three branches—classification, regression, and segmentation—by incorporating a fully convolutional network, which significantly enhances detection accuracy. The algorithm follows a two-stage framework: the first stage uses a residual network, a feature pyramid network, and a region proposal network to generate candidate bounding boxes; the second stage employs ROI Align to map these regions to the feature map, followed by classification, bounding box regression, and segmentation *(Ren et al., 2022; Zhang et al., 2020; Zhang et al., 2022)*.

In this study, the standard Feature Pyramid Network (FPN) was modified by incorporating the bidirectional fusion mechanism from EfficientDet, creating a new BF-FPN (Bidirectional Fusion-FPN). Additionally, a CPCA module was introduced after the ResNet backbone and added a self-attention mechanism in the mask branch. To further improve boundary segmentation accuracy, the DiceLoss function was included in the mask loss calculation. The improved network structure is shown in Figure 2.

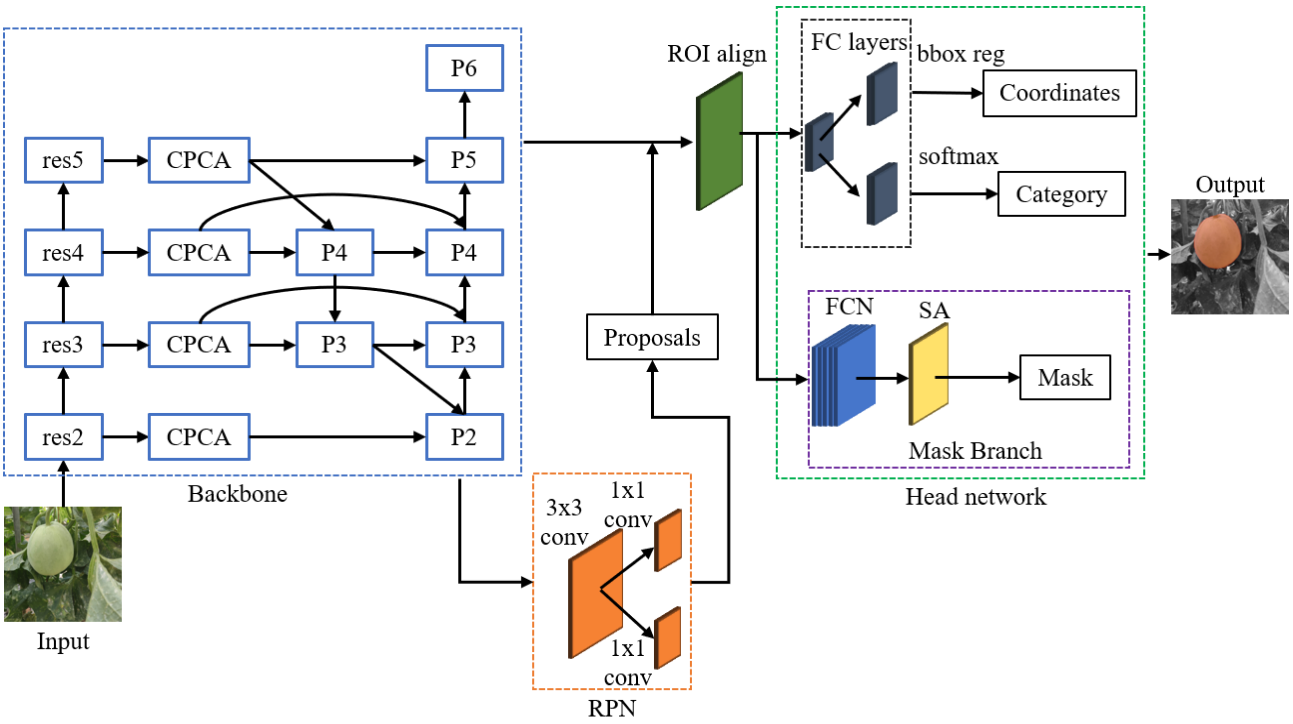


Fig. 2 - Improved Mask R-CNN Network Structure Diagram

1 - Backbone is selected as ResNet50; 2 - conv refers to convolution;
3 - CPCA represents channel prior attention mechanism;
4 - RPN refers to region proposal network; 5 - FC layers refer to fully connected layers;
6 - FCN refers to fully connected network; 7 - SA refers to self-attention mechanism.

CPCA Attention Mechanism

Attention mechanisms in deep learning simulate the human cognitive ability to focus selectively on important information, dynamically assigning weights to highlight relevant features and suppress irrelevant ones, thereby improving detection accuracy *(Wu et.al., 2025)*. The Channel Prior Convolutional Attention (CPCA) mechanism dynamically allocates attention weights across channel and spatial dimensions. It extracts multi-scale spatial information via depthwise separable convolutions and generates spatial attention maps that better reflect real feature distributions, significantly enhancing segmentation performance. CPCA combines the strengths of both channel and spatial attention, making it a highly efficient attention mechanism *(Huang et al., 2024; Liu et al., 2025)*. Its structure is illustrated in Figure 3.

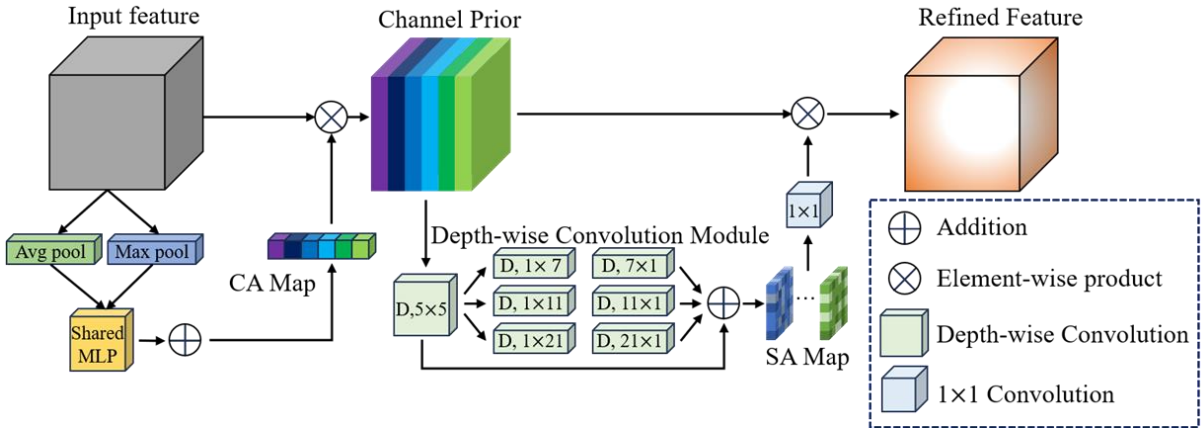


Fig. 3 - CPCA Structure Diagram

Channel Attention: Applies average pooling and max pooling to the input, aggregates spatial information from the feature maps, and feeds them into a shared MLP followed by a Sigmoid activation to produce attention-weighted features.

$$CA(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) \quad (1)$$

wherein: $\boldsymbol{\sigma}$ denotes Sigmoid function.

Spatial Attention: Captures spatial relationships using depthwise separable convolutions and enhances this with a multi-branch structure to improve spatial representation while reducing computational load.

$$SA(F) = Conv_{1\times1}\left(\Sigma_{i=0}^{3} Branch_i\left(DwConv(F)\right)\right) \quad (2)$$

wherein: $\boldsymbol{Dwconv}$ denotes depthwise convolution, and $\boldsymbol{Branch_i}$ represents the $\boldsymbol{i}$ - th branch.

Improved FPN Network

Feature Pyramid Network (FPN) is widely used to extract multi-scale features by combining semantically rich high-level features with fine-grained low-level features *(Lin et al., 2017)*. Traditional FPN utilize a top-down pathway to propagate high-level features downward; however, this unidirectional fusion may result in insufficient detail preservation and poor scale balance. To address this limitation, the bidirectional fusion mechanism from EfficientDet was adopted to restructure the original FPN *(Jeon et al., 2022)*. The modifications include output-stage downsampling, removal of nodes with low feature contributions, and the addition of lateral connections between Res3 and P3, as well as between Res4 and P4. The improved Bidirectional Fusion FPN (BF-FPN) achieves better balance in multi-scale feature representation and enhances feature learning across different levels. The network architectures before and after the modification are illustrated in Figure 4.

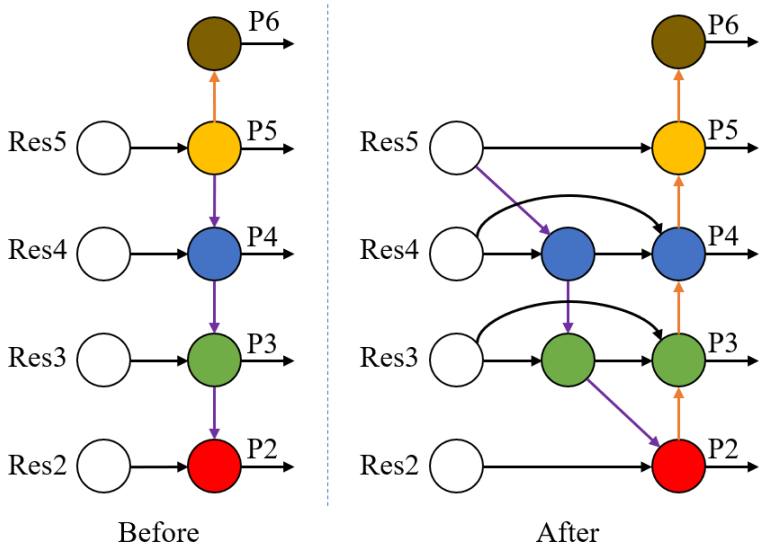


Fig. 4 - Comparison Diagram of FPN Before and After Improvement

The improved backbone network incorporates the BF-FPN and the CPCA attention mechanism to enhance feature extraction capabilities, thereby increasing object detection accuracy and segmentation precision.

Improved Mask Branch

The self-attention mechanism dynamically adjusts the weight of each pixel based on its relationship with all other pixels in the input feature map. This enables the model to focus more effectively on relevant regions during the segmentation process, particularly when dealing with complex object shapes or indistinct boundaries *(Li et al., 2020)*.

To improve mask segmentation accuracy, a self-attention (SA) module was integrated into the mask branch of the Mask R-CNN. The module first generates query, key, and value feature maps using three separate 1×1 convolutional layers. It then computes the attention map by performing dot-product operations between the query and key maps, followed by Softmax normalization. Finally, a weighted sum of the value map is calculated based on the attention scores. This mechanism allows the output features at each pixel to incorporate both local and global contextual information more effectively. The structure and insertion point of the SA module are shown in Figure 5.

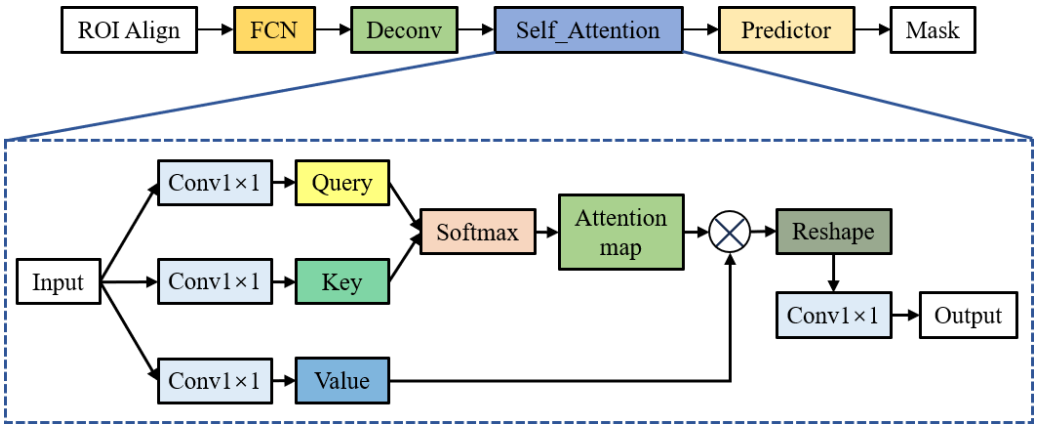


Fig. 5 - Insertion Position and Structural Diagram of SA

The loss function, also known as the cost function, is used to evaluate the degree of difference between the predicted and actual values of a model. The training process is the process of minimizing the loss function. The smaller the loss function, the closer the predicted value of the model is to the true label, indicating that the robustness of the model is better (*Li J. et al, 2023*). To further improve boundary segmentation accuracy, Dice Loss was added to the original mask loss function. The Dice Loss function is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2} \quad (3)$$

where $\boldsymbol{N}$ is the total number of pixels, $\boldsymbol{p_i}$ represents the predicted probability for the $\boldsymbol{i}$-th pixel, and $\boldsymbol{g_i}$ is the corresponding ground truth label (0 or 1).

The modified mask loss function is expressed as:

$$L_{mask} = \alpha L_{BCE} + (1-\alpha) L_{Dice} \quad (4)$$

where the weight $\boldsymbol{\alpha}$ was experimentally set to 0.3 for optimal performance. $\boldsymbol{L_{BCE}}$ is the binary cross-entropy loss of the original algorithm.

Measurement Method

In the melon diameter measurement system, a binocular camera captures stereo images of melons via a USB connection. The camera has a total resolution of 2560 × 720 pixels and a baseline distance of 4 cm. A portable computer processes the captured images and outputs the diameter measurements. The software is implemented in Python, utilizing libraries such as NumPy and OpenCV for image acquisition, transmission, computation, and display. The experimental setup is illustrated in Figure 6.

Fig. 6 - Experimental environment

To obtain depth information from the binocular camera, camera calibration, stereo rectification, and stereo matching must be performed. First, Zhang Zhengyou calibration method is applied to calibrate the binocular camera, yielding the intrinsic and extrinsic parameters of both cameras *(Li, 2020; Zhang et al., 2024)*. The selected 2D image coordinate system is assumed to be parallel to the 3D world coordinate plane at Z=0. The coordinate system of the left camera is defined as the world coordinate system in the binocular vision measurement setup. For a spatial point M(x, y, z), its corresponding image projection point on the left camera M1 has pixel coordinates (u, v). The transformation relationship between M and M1 is given as follows:

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (5)$$

where Z represents the depth of the point in 3D space; f_x and f_y re the focal lengths of the camera in the x and y directions, respectively; (u_0, v_0) is the principal point; R and T are the external parameters representing the rotation matrix and translation vector, respectively.

Next, stereo rectification is applied to align the image pairs onto the same plane. After correcting for lens distortion, the rectified left and right images are processed using row alignment to reduce the matching search from 2D to 1D space.

The Semi-Global Block Matching (SGBM) algorithm was adopted to compute the disparity map, which estimates the pixel displacement d between corresponding points in the left and right images. Given the camera baseline B and focal length f, the depth Z is calculated using:

$$Z = \frac{f \times B}{d} \quad (6)$$

Based on accurately segmented melon contours, the system identifies key measurement points. First, the upper half of the melon region is sampled for curvature analysis. Among the sampled points, the point with the maximum curvature is selected as the vertical diameter point L1. The curvature is calculated using three consecutive points at a time. For example, given three points A, B, and C that form a triangle with side lengths a, b, and c, the area of the triangle S_Δ is computed as follows:

$$S_\Delta = \sqrt{p(p-a)(p-b)(p-c)} \quad (7)$$

where $p = (a + b + c) / 2$.

Then, the curvature K at point B is calculated based on the geometry of triangle ABC, using the following formula:

$$K = \frac{4\sqrt{p(p-a)(p-b)(p-c)}}{abc} \quad (8)$$

The point L2 is determined by extending the line from L1 to the center of the minimum bounding rectangle (MBR) until it intersects with the lower half of the contour. This extended line defines the vertical axis L. Then, a perpendicular line T to L is drawn, and its intersection points with the contour define the horizontal diameter points T1 and T2, selected based on maximum distance. The measurement point selection process is illustrated in Figure 7.

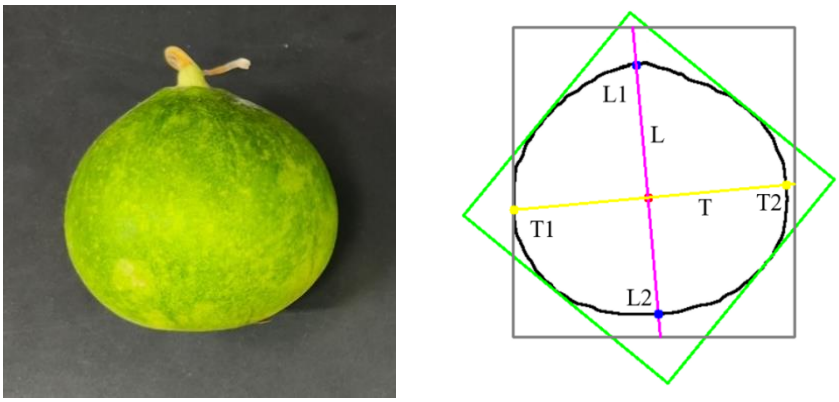


Fig. 7 - Schematic diagram of measurement point selection

After obtaining the pixel coordinates of the four measurement points, their depth values are acquired using the binocular camera system. These depth values are then used to convert the pixel coordinates into spatial (3D) coordinates. Finally, the Euclidean distance between the corresponding points for the horizontal and vertical diameters is calculated to determine the actual fruit dimensions.

Evaluation Metrics

To validate the effectiveness of the improved Mask R-CNN, both the detection and segmentation performance were evaluated using precision (P), recall (R), and the F1-score *(Wang et al., 2024; Zhao et al., 2022)*. The definitions of these evaluation metrics are as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (9)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (10)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (11)$$

where: $\boldsymbol{TP}$ (True Positive): predicted as positive and actually positive. $\boldsymbol{FP}$ (False Positive): predicted as positive but actually negative. $\boldsymbol{FN}$ (False Negative): predicted as negative but actually positive.

To evaluate the accuracy of the proposed measurement system, the relative error ($\boldsymbol{E_R}$) between the system-measured and manually measured fruit diameters was calculated using the following formula:

$$E_R = \frac{|L_s - L_m|}{L_t} \times 100\% \quad (12)$$

where $\boldsymbol{L_s}$ represents the system-measured diameter and $\boldsymbol{L_m}$ represents the manually measured diameter. To assess the overall accuracy across multiple samples, the Mean Relative Error (***MRE***) was computed as:

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|L_s^{(i)} - L_m^{(i)}\right|}{L_m^{(i)}} \times 100\%\right) \quad (13)$$

where: $\boldsymbol{n}$ is the total number of samples, and $\boldsymbol{L_s^{(i)}}$ and $\boldsymbol{L_m^{(i)}}$ are the system and manual measurements of the $\boldsymbol{i}$ – th sample, respectively. A lower MRE indicates higher measurement accuracy and reliability.

RESULTS

Ablation experiment

To evaluate the impact of each module and analyze the performance of the improved algorithm, ablation experiments were conducted on the self-built melon dataset. Precision (P) and F1-score were used as evaluation metrics, where higher values indicate better algorithm performance. The effects of each module on detection and segmentation results are summarized in Table 1.

Table 1

Ablation test results for different improvement points

Model	Detection			Segmentation		
	P%	R%	F1%	P%	R%	F1%
Mask R-CNN	91.3	92.8	92.0	90.5	91.8	91.1
+ BF-FPN	93.5	93.6	93.5	91.7	92.3	92.0
+ BF-FPN + CPCA	94.0	93.3	93.6	92.3	92.7	92.5
+ SA	91.2	92.6	91.9	90.8	92.1	91.4
+ SA + Dice Loss	91.4	92.7	92.0	91.1	92.7	91.9
+ BF-FPN + CPCA + SA + Dice Loss	94.2	94.5	94.3	92.7	93.6	93.1

As shown in the table, the introduction of the BF-FPN module led to improvements in both detection and segmentation precision, by 2.2% and 1.2% respectively, with corresponding increases in the F1-score of 1.5% and 0.9%. Building upon this, the addition of the CPCA module between ResNet and BF-FPN further enhanced detection and segmentation precision by 0.5% and 0.6%, respectively. Subsequently, a self-attention (SA) module was integrated into the mask branch to improve mask segmentation accuracy, resulting in a 0.5% increase compared to the baseline. Incorporating the DiceLoss function into the loss calculation raised the segmentation accuracy to 91.1%, which represents a 0.6% improvement over the original model. Finally, by combining the improved detection and segmentation modules, the overall detection and segmentation precision increased by 2.9% and 2.2%, recall improved by 1.7% and 1.8%, the F1-score rose by 2.3% and 2.0% compared to the baseline.

Comparative Analysis with Other Algorithms

The improved Mask R-CNN was compared with other mainstream instance segmentation models, including PointRend, TensorMask, Cascade Mask R-CNN, YOLACT, and SOLOv2. The results are shown in Table 2.

Table 2

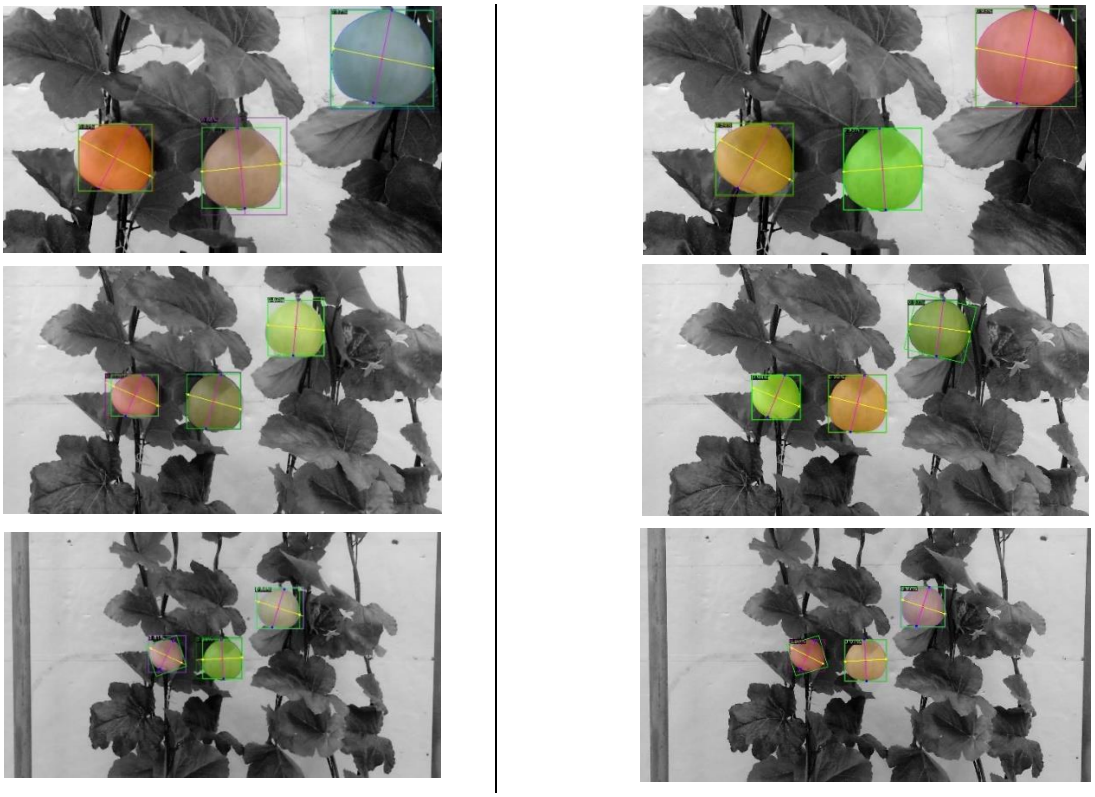
Comparison of detection effects of different models

Model	Detection			Segmentation		
	P%	R%	F1%	P%	R%	F1%
PointRend	85.7	87.3	86.5	79.4	81.7	80.5
Tensormask	86.6	92.5	89.9	85.2	89.4	87.3
Casced Mask R-CNN	91.7	93.6	92.6	90.8	90.2	90.5
YOLACT	90.6	87.1	88.8	86.4	89.3	87.8
SOLOv2	91.2	86.8	88.9	89.2	91.7	90.4
Improved	94.2	94.5	94.3	92.7	93.6	93.1

Compared with the best alternatives, the improved model outperforms in both detection and segmentation metrics. Compared with PointRend and TensorMask, it offers up to 8.5% and 7.6% improvements in detection accuracy, and 13.3% and 7.5% improvements in segmentation accuracy, respectively. This confirms the superiority of the proposed method for precise melon fruit detection and segmentation.

Fruit Diameter Measurement Experiment and Analysis

Three melon samples were selected for the experiment. A vernier caliper was used to measure both the horizontal and vertical diameters three times, and the average values were taken as the manual measurements. A fruit diameter measurement experiment was conducted using the HBVCAM binocular camera. The camera was mounted on a vertical frame at a height of 0.5 meters from the ground, with the shooting angle parallel to the ground. Images were captured at distances of 0.3 m, 0.5 m, and 0.7 m. From left to right in the figure, the melons are labeled as Melon1, Melon2, and Melon3. The proposed algorithm was used to identify the melons and locate the measurement points for the vertical (V) and horizontal (H) diameters. The actual fruit diameters were then calculated using stereo vision. The measurement point localization results before and after algorithm improvement are shown in Figure 8.



Original Improved

Fig. 8 - Comparison of Measurement Point Localization Effects Between Original Algorithm and Improved Algorithm

Measurement results are summarized in the following Table 3:

Table 3

Result of measurement

Distance (m)	Melon ID	Manual (mm)		System (mm)		E_R (%)	
		V	H	V	H	V	H
0.3	Melon1	55.2	67.5	50.5	61.2	8.5	9.3
	Melon2	81.1	80.9	74.7	74.3	7.9	8.2
	Melon3	76.4	87.1	70.5	81.7	7.7	6.2
0.5	Melon1	55.2	67.5	51.3	62.4	7.1	7.6
	Melon2	81.1	80.9	75.2	74.7	7.3	7.7
	Melon3	76.4	87.1	71.2	80.5	6.8	7.6
0.7	Melon1	55.2	67.5	59.2	73.1	7.2	8.3
	Melon2	81.1	80.9	74.5	87.7	8.1	8.4
	Melon3	76.4	87.1	82.4	93.6	7.9	7.5

The results demonstrate that the proposed method achieves MRE of 7.1% for horizontal diameters and 7.6% for vertical diameters, indicating a satisfactory level of accuracy for practical applications.

System Interface Design

The melon fruit diameter measurement application was developed using PyQt5. PyQt5 offers rich GUI functionalities, strong extensibility, and high integration with Python, the programming language used in this research. Therefore, it is well suited for implementing the proposed method, as shown in Figure 9.

The program mainly implements the following functions:

1） Real-time Image Transmission: The application can acquire video captured by the camera and display it in the "Camera View" area.

2） Image Processing Result Display: Processed information is displayed on the original image as well as in the "Detection Result" area.

3） Fruit Diameter Measurement Result Display: The measured fruit diameter results are outputted and displayed in the "Measurement Output" area.

4） Control Functions: Buttons are designed to control the system's start and stop of recognition, as well as the saving of results.

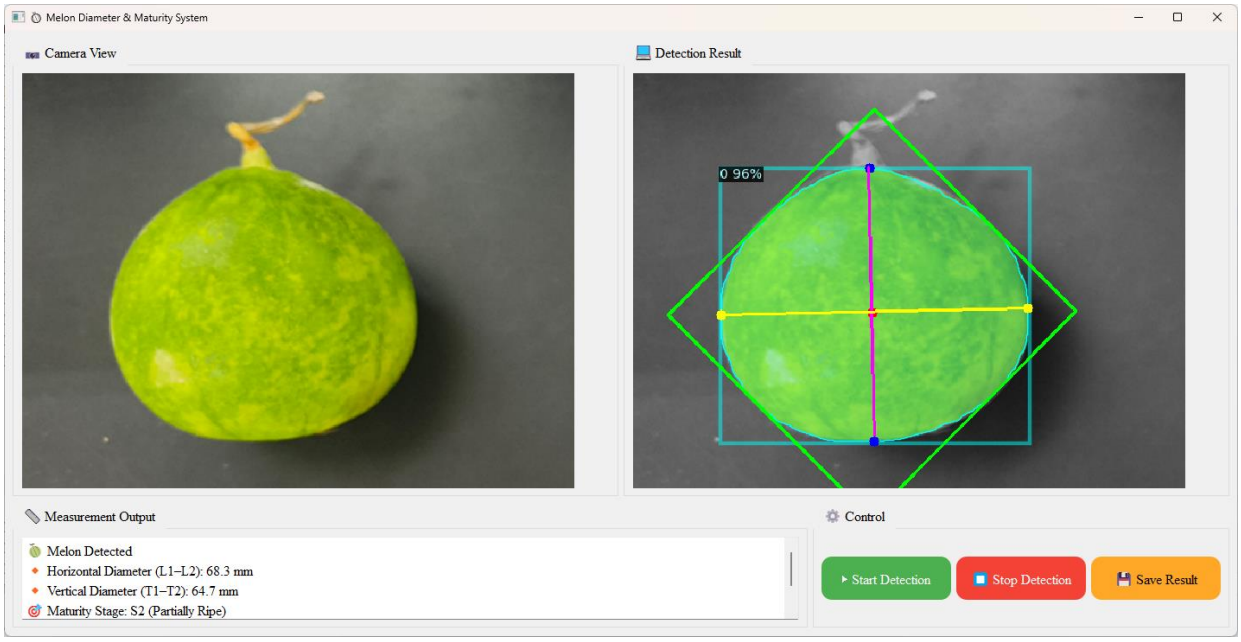


Fig. 9 – The Measurement System

CONCLUSIONS

This paper presents a method for measuring melon fruit diameter based on image segmentation and binocular vision. Image data are captured using a binocular camera, and melons are detected and segmented with an improved Mask R-CNN algorithm. Measurement points on the fruit surface are identified by analyzing curvature and other geometric features. Depth information for these points is then obtained by mapping 2D image coordinates to 3D space. Finally, the Euclidean distance is used to calculate the fruit diameter.

(1) Based on Mask R-CNN, this study introduces a newly designed BF-FPN and incorporates the CPCA attention mechanism at its input to enhance the algorithm's ability to detect targets of varying sizes. Additionally, a self-attention mechanism is integrated into the mask branch to improve edge segmentation accuracy. As a result, the F1 scores for object detection and segmentation increased by 2.3% and 2%, respectively.

(2) By refining the region of interest within the detection box, the candidate area for measurement point selection is narrowed. Curvature analysis of the melon contour is performed, and the minimum enclosing rectangle is fitted to determine the pixel coordinates of key measurement points. Their depth values are calculated using stereo vision, and the actual fruit diameter is obtained via Euclidean distance computation. Experimental results demonstrate that the proposed method achieves mean relative errors of 7.1% for horizontal diameter and 7.6% for vertical diameter, satisfying practical accuracy requirements.

(3) This study offers a non-contact approach to melon diameter measurement, improving operational efficiency and providing a methodological reference for assessing fruit ripeness.

ACKNOWLEDGEMENT

This research was supported by the Natural Science Foundation of Shandong Province, grant number [ZR2022MF306] and Shandong Province Agricultural Major Application Technology Innovation Project (SD2019NJ001). The APC was funded by Xincheng Li.

REFERENCES

[1] Basak, J. K., Paudel, B., Kim, N. E., Deb, N. C., Madhavi, B. G. K., & Kim, H. T. (2022). Non-destructive estimation of fruit weight of strawberry using machine learning models. *Agronomy*, *12*(10), 2487. https://doi.org/10.3390/agronomy12102487

[2] Chang, L.-Y., He, S.-P., Liu, Q., Xiang, J.-L., & Huang, D.-F. (2018). Quantifying muskmelon fruit attributes with A-TEP-based model and machine vision measurement. *Journal of Integrative Agriculture*, *17*(6), 1369–1379. https://doi.org/10.1016/S2095-3119(18)61912-4

[3] Geng, A., Gao, A., Yong, C., Zhang, Z., Zhang, J., & Zheng, J. (2022). Dropping ear detection method for corn harvester based on improved Mask-RCNN. *INMATEH – Agricultural Engineering*, *66*(1), 31–40. https://doi.org/10.35633/inmateh-66-03

[4] Gothi, H. R., Patel, P. S., Raj, V. P., Rabari, P. H., Balina, P. K., Sharma, S. K., & Ghetiya, L. V. (2022). Diversity and abundance of insect pollinators on muskmelon. *Journal of Entomological Research*, *46*(Suppl.), 1102–1107. https://doi.org/10.5958/0974-4576.2022.00187.6

[5] Gu, Z., Ma, X., Guan, H., Jiang, Q., Deng, H., Wen, B., Zhu, T., & Wu, X. (2024). Tomato fruit detection and phenotype calculation method based on the improved RTDETR model. *Computers and Electronics in Agriculture*, *227*(Part 1), 109524. https://doi.org/10.1016/j.compag.2024.109524

[6] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. *arXiv preprint* arXiv:1703.06870. https://arxiv.org/abs/1703.06870

[7] Huang, H., Chen, Z., Zou, Y., Lu, M., Chen, C., Song, Y., Zhang, H., & Yan, F. (2024). Channel prior convolutional attention for medical image segmentation. *Computers in Biology and Medicine*, *178*, 108784. https://doi.org/10.1016/j.compbiomed.2024.108784

[8] Jeon, K. J., Ha, E.-G., Choi, H., Lee, C., & Han, S.-S. (2022). Performance comparison of three deep learning models for impacted mesiodens detection on periapical radiographs. *Scientific Reports*, *12*(1), 15402. https://doi.org/10.1038/s41598-022-19753-w

[9] Li, H., & Tang, J. (2020). Dairy goat image generation based on improved self-attention generative adversarial networks. *IEEE Access*, *8*, 62448–62457. https://doi.org/10.1109/ACCESS.2020.2981496

[10] Li, J., Liu, K., Hu, Y., Zhang, H., Heidari, A. A., Chen, H., Zhang, W., Algarni, A. D., & Elmannai, H. (2023). Eres-UNet++: Liver CT image segmentation based on high-efficiency channel attention and Res-UNet++. *Computers in Biology and Medicine*, *158*, 106501. https://doi.org/10.1016/j.compbiomed.2022.106501

[11] Li, Y. (2020). A calibration method of computer vision system based on dual attention mechanism. *Image and Vision Computing*, *103*, 104039. https://doi.org/10.1016/j.imavis.2020.104039

[12] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *arXiv preprint* arXiv:1612.03144. https://arxiv.org/abs/1612.03144

[13] Liu, H., Jiang, Y., Zhang, W., Li, Y., & Ma, W. (2025). Intelligent electronic components waste detection in complex occlusion environments based on the focusing dynamic channel–you only look once model. *Journal of Cleaner Production*, *486*, 144425. https://doi.org/10.1016/j.jclepro.2024.144425

[14] Ren, X., Sun, M., Zhang, X., Liu, L., Zhou, H., & Ren, X. (2022). An improved Mask R-CNN algorithm for UAV TIR video stream target detection. *International Journal of Applied Earth Observation and Geoinformation*, *106*, 102660. https://doi.org/10.1016/j.jag.2021.102660

[15] Wang, N., Li, X., Shang, S., Yun, Y., Liu, Z., & Lyu, D. (2024). Monitoring dairy cow rumination behavior based on upper and lower jaw tracking. *Agriculture*, *14*(11), 2006. https://doi.org/10.3390/agriculture14112006

[16] Wang, Y., Wen, L., Zhao, D., Wang, G., Jia, Y., & Su, X. (2024). Long-season cultivation techniques for thin-skinned muskmelon. *Northern Horticulture*, (13), 156–158. https://doi.org/CNKI:SUN:BFYY.0.2024-13-023

[17] Wu, X., Ma, Y., Zhang, S., Chen, T., & Jiang, H. (2025). Yo3RL-Net: A fusion of two-phase end-to-end deep net framework for hand detection and gesture recognition. *Alexandria Engineering Journal*, *121*, 77–89. https://doi.org/10.1016/j.aej.2025.01.097

[18] Xue, Q., Li, H., Chen, J., & Du, T. (2024). Fruit cracking in muskmelon: Fruit growth and biomechanical properties under different irrigation levels. *Agricultural Water Management*, *293*, 108672. https://doi.org/10.1016/j.agwat.2024.108672

[19] Zhang, Q., Chang, X., & Bian, S. (2020). Vehicle-damage-detection segmentation algorithm based on improved Mask R-CNN. *IEEE Access*, *8*, 6997–7004. https://doi.org/10.1109/ACCESS.2020.2964055

[20] Zhang, Q., & Tang, F. (2024). Implementation of drill pipe joint positioning based on binocular vision. *Petroleum Machinery*, *52*(10), 12–19+73. https://doi.org/10.16082/j.cnki.issn.1001-4578.2024.10.002Zhang, R., Jia, Z., Wang, R., Yao, S., & Zhang, J. (2022). Phenotypic parameter extraction for wheat ears based on an improved Mask-RCNN algorithm. *INMATEH–Agricultural Engineering*, *66*(1), 267–278. https://doi.org/10.35633/inmateh-66-27

[21] Zhao, F., Zhang, J., Zhang, N., Tan, Z., Xie, Y., Zhang, S., Han, Z., & Li, M. (2022). Detection of cucurbits' fruits based on deep learning. *INMATEH–Agricultural Engineering*, *66*(1), 321–330. https://doi.org/10.35633/inmateh-66-32

[22] Zheng, L., Wang, L., Wang, M., & Ji, R. (2021). Automated 3D point cloud reconstruction of oilseed lettuce based on Kinect camera. *Transactions of the Chinese Society for Agricultural Machinery*, *52*(7), 159–168.