Research on Target Recognition, Segmentation, and Accurate Positioning Methods for Obscured Apple Fruits Under Dense Canopies

Jiang Houkang[a], Liu Jizhan[a,b,c*],Lei Xiaojie[a], Xu Baocheng[a], Wang Jie[a]

[a] *Key Laboratory for Theory and Technology of Intelligent Agricultural Machinery and Equipment, Jiangsu University, Zhenjiang, China*

[b.] *Jiangsu Province and Education Ministry Co-sponsored Synergistic Innovation Center of Modern Agricultural Equipment, Jiangsu University, Zhenjiang, China*

[c] *National Digital Agricultural Equipment (Artificial Intelligence and Agricultural Robotics) Innovation Sub-Centre, Jiangsu University, Zhenjiang, China*

ABSTRACT

In robotic harvesting, recognizing and locating target apple fruits, as well as detecting obstacles in complex environments like dense canopies, poses significant challenges. This paper presents a method for recognizing, segmenting, and accurately localizing obscured fruits. We employ instance segmentation to extract edge contours of fruits and branches, leveraging two-dimensional features for precise localization under occlusion and correcting picking point errors. Experimental results indicate that at a 10% occlusion ratio, the model achieves an average precision of 96.8%, an F1 Score for edge detection of 0.91, and a 3D reconstruction error of 3.1 mm. Even at 50% occlusion, average precision remains at 88.3%, with an F1 Score of 0.80 and a reconstruction error of 7.6 mm. These findings demonstrate that while occlusion significantly impacts 3D reconstruction, the algorithm retains high accuracy, fulfilling the precision requirements for fruit harvesting.

Keywords: fruit recognition, edge detection, robotic harvesting, 3D reconstruction, precise localization.

1. INTRODUCTION

Against the backdrop of agricultural labor shortages, robotic harvesting has become a focal point of research, offering an alternative to traditional manual picking methods [1]. A variety of harvesting robots have already been developed and implemented in the market, including apple picking robots [2-4], grape harvesting robots [5-6], and kiwi picking robots [7-8]. Supported by deep learning image processing techniques and depth information 3D positioning technologies, the accuracy of visual systems in harvesting robots has significantly improved [9]. In 2021, Chu et al. proposed a method to suppress the Mask R-CNN model for the recognition of two apple varieties, "Gala" and "Blondee," in natural environments, meeting the real-time requirements of apple picking robots [10]. In 2019, Yang et al. addressed the strawberry picking problem in unstructured environments by utilizing Mask R-CNN to recognize various strawberry varieties such as "Hongyan" and "Zhajie" and calculate picking points [11]. In 2017, Liu Jizhan et al. studied citrus fruits, leaves, and branches, each exhibiting distinct three-dimensional geometric features such as spheres, planes, and cylindrical shapes [12]. However, in complex environments like orchards with dense canopies, deep learning techniques and traditional machine vision approaches can only provide fruit location information without simultaneously acquiring details about surrounding obstacles. This limitation prevents the correct identification of picking points for occluded fruits, failing to address issues caused by point occlusion and overlapping fruits in orchard operations. Furthermore, improper picking positions may lead to fruit damage and even harm the robotic arm, resulting in a chaotic workflow that severely affects operational efficiency. To address the shortcomings of existing technologies, we propose a method for the recognition, segmentation, and precise localization of obscured fruit targets under dense canopies, along with feasibility planning for harvesting. This method utilizes instance segmentation to extract edge contours of fruits and branches, combined with the two-dimensional features of fruit spheres, enabling precise localization under occlusion and correcting errors in three-dimensional picking point localization, thereby guiding the robotic arm to harvest fruits effectively.

International Conference on Optical and Photonic Engineering (icOPEN 2024), edited by Xiangjun Zou,
Jianglei Di, Kemao Qian, Shijie Feng, Jianping Zhou, Proc. of SPIE Vol. 13509, 135090U
© 2025 SPIE · 0277-786X · doi: 10.1117/12.3057765

2. METHOD FOR TARGET RECOGNITION AND SEGMENTATION OF OBSCURED FRUITS UNDER DENSE CANOPIES

2.1 Improved Mask R-CNN Fruit Target Recognition and Segmentation Method

2.1.1 Mask R-CNN Instance Segmentation Method

Mask R-CNN is a novel object detection framework introduced in 2017. Compared to models such as Faster R-CNN, YOLO, and SSD, Mask R-CNN achieves high-quality instance segmentation while effectively recognizing targets by generating masks with customizable shapes.

Mask R-CNN primarily consists of several components: image input, ResNet, Feature Pyramid Network (FPN), Region Proposal Network (RPN), ROIAlign, and image output (bounding boxes, classes, and masks). A simplified diagram is shown in Figure 1.

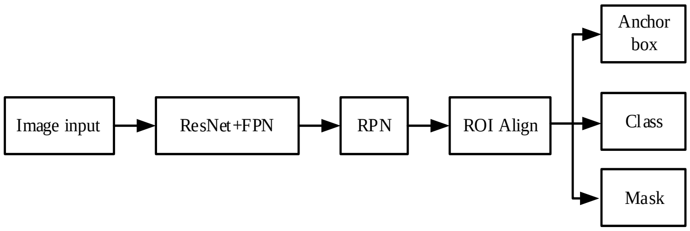

Figure 1. Simplified Diagram of the Mask R-CNN Network.

2.1.2 Fruit Multi-Scale Processing and Fusion Method Based on FPN and PANet

The diversity in fruit morphology, size, and color, along with overlapping and occlusion among apple fruits, leads to unclear boundaries and difficulties in segmentation. To address these challenges, we integrate Feature Pyramid Networks (FPN) and Path Aggregation Networks (PANet) into the improved Mask R-CNN framework for effective target detection and recognition.

FPN constructs a multi-scale feature pyramid using a bottom-up fusion mechanism that combines feature maps of varying resolutions, allowing better capture of detailed image information. This is crucial for small fruits hidden among foliage, requiring precise localization on high-resolution maps. The feature pyramids are denoted as P_2, P_3, P_4, P_5, with P_2 representing the highest resolution layer. The top-down structure of FPN can be expressed as:

$$P_l = Conv(P_l + 1) + Upsample(P_l + 1), l = 2,3,4 \tag{1}$$

where $Conv(.)$ represents the convolution operation, and $Upsample(.)$denotes the upsampling operation.

For occluded and partially visible apple fruits, the issues of unclear boundaries and difficulty in observing and segmenting certain fruit parts can be addressed by combining FPN with PANet, which utilizes multi-layer feature fusion.This allows the model to achieve more accurate target localization and boundary delineation, supported by high-level semantic information. PANet introduces additional path enhancement modules that re-integrate low-level feature information into high-level semantic features through bottom-up feature fusion, thus improving the model's ability to handle ambiguous boundary targets such as occluded and incomplete fruits. By increasing bottom-up paths, it supplements high-level semantic information within the FPN structure, enriching the final feature representation. In PA

$$P_l{}' = P_l + Conv(N_l)\ , l = 2,3,4,5 \tag{2}$$

where $P_l{}'$ denotes the enhanced feature maps, and $Conv(N_l)$ represents the convolution operation applied to low-level features.

Through multi-layer feature aggregation, the model can effectively utilize spatial details and semantic information, enabling accurate segmentation of boundaries for occluded and partially visible apple fruits.

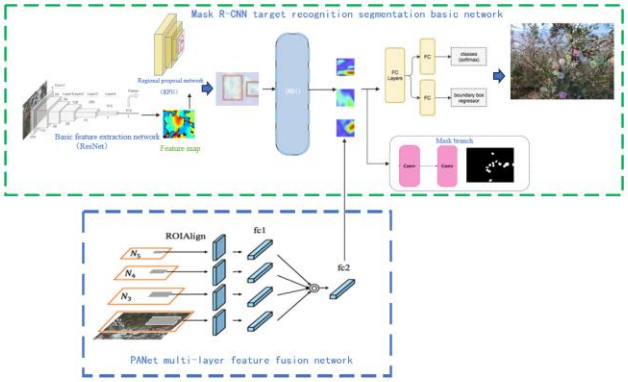
Mask R-CNN target recognition segmentation basic network
Regional proposal network
(RPN)
Basic feature extraction network
(ResNet)
Feature map
ROI
FC Layers
FC
FC
classes (softmax)
boundary box regressor
Mask branch
Conv
Conv
ROIAlign
fc1
fc2
N_5
N_4
N_3
N_2
PANet multi-layer feature fusion network

Figure 2. Architecture of the Fusion Mask R-CNN Network Based on FPN and PANet.

2.1.3Model Loss Function

In the improved Mask R-CNN model, the loss function design is essential for enhancing accuracy and robustness. Given the diversity in fruit shapes, sizes, and colors, as well as occlusions and overlaps, the loss function must address classification, localization, boundary delineation, and instance segmentation.The classification loss measures the accuracy of the model for candidate regions (ROIs). Mask R-CNN typically uses a multi-class cross-entropy loss, represented as:

$$L_{cls} = -\sum_{i=1}^{N} y_i \log(p_i) \tag{3}$$

where y_i is the ground truth label, p_i is the predicted probability of the class, and N is the total number of classes.

To improve the localization accuracy of the fruits, Mask R-CNN uses the Smooth L1 loss to measure the error between the predicted bounding box and the ground truth bounding box. Let the ground truth bounding box be t^*， and the predicted bounding box be t， then the bounding box regression loss is represented as:：

$$L_{bbox} = -\frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} smooth_{L_1}(t_i - t_i^*)] \tag{4}$$

where N_{pos} is the number of positive samples, and $smooth_{L_1}(.)$is the Smooth L_1 loss function, which helps avoid gradient explosion caused by outliers while ensuring accuracy.

To accurately segment the shape and boundaries of the fruits, the mask loss employs binary cross-entropy loss, which calculates the error between the predicted mask and the ground truth mask for each pixel location:

$$L_{mask} = -\frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{5}$$

By minimizing this loss, the model can more accurately segment the fruit instances.

Finally, the total loss function is a weighted sum of the three losses, expressed as:

$$L_{total} = L_{cls} + \lambda_1 L_{bbox} + \lambda_2 L_{mask} \tag{6}$$

where λ_1 andλ_2 are the weights for regulating the bounding box regression loss and the mask loss, respectively. By adjusting the weight of these two parameters, a balance can be achieved between classification, localization, and segmentation.

2.1.4 Creation and Training of the Fruit Dataset

The images and depth point clouds used in this study were collected at the 10,000-acre orchard of Siweite Fruit Co., Ltd. in Suqian, Jiangsu Province, China, on November 21, 2022, from 10:30–11:30 AM and 2:30–5:30 PM. The collection was conducted using a RealSense D435 depth camera with the RealSenseViewer V2.55.1 SDK, capturing synchronized 640×480 resolution, 8-bit color channel data. To address scale variations, the collectors took images at distances of 0.8m, 1.2m, and 1.6m from each fruit tree along the rows to ensure clarity and size consistency of the fruit in the images.

The collected images were annotated using the LabelImg software. The annotation method involved enclosing each apple with the smallest possible curved polygon to capture both the apple's outer contour and its central coordinates. The annotation process is shown in Figure 3.

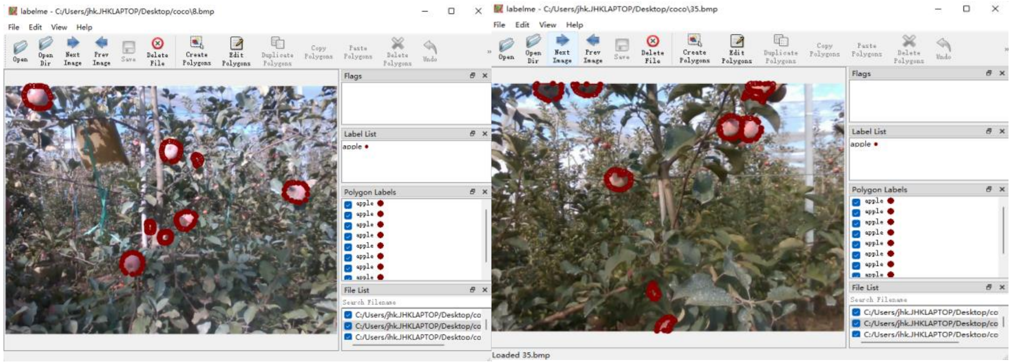

Figure 3. LabelImg Annotation Software Interface.

Before configuring and training the model, the annotated dataset was divided into a training set and a test set, with 2000 and 1000 images, respectively. The network used is mask_rcnn_R_50_FPN_3x (backbone: ResNet-50 + Feature Pyramid Network, FPN).The parameter settings involved in training the model for this experiment are shown in Table 1:

Table 1. Model Training Parameters

Parameter	Value
Learning rate	0.002
Momentume	0.910
Decay Strategy	cosine
Max Iterations	12000
Batch size	16
Weight-decay	0.0001

3. OCCLUDED FRUIT COMPLETION AND PRECISE LOCALIZATION METHOD

3.1 2D Completion Method for Touching and Occluded Fruits

3.1.1 Pixel-Level Segmentation and Contour Extraction of Apple Fruits

Given the complex lighting conditions in orchards, fruit color and brightness may vary significantly across different areas, making it challenging to consistently identify fruit boundaries. Fruits are often closely packed or in contact with each other, partially or entirely occluded by branches and leaves. To accurately delineate pixel-level fruit boundaries for completing occluded fruits and achieving precise localization, a region proposal network (RPN) is used to generate candidate regions. These candidate regions are represented by bounding boxes, each containing potential fruit targets. The bounding boxes generated by the RPN can be represented by four coordinates$(x_{min}, y_{min}, x_{max}, y_{max})$, defining the spatial scope of each ROI. For each ROI, the mask head uses a Fully Convolutional Network (FCN) to generate a binary mask that captures the pixel-level boundary of the fruit:

$$M = FCN(F_{roi}) \tag{7}$$

where M is the generated binary mask matrix, with each pixel value indicating whether it belongs to the target fruit.

Through the multi-scale feature fusion of FPN and PANet, the mask M is refined by combining feature maps from different levels to better capture the boundary details of the fruits. The fused feature map is represented as F_{pan}, and the refined mask is expressed as:

$$M' = Refine(M, F_{pan}) \tag{8}$$

where $Refine(.)$ denotes the mask refinement process. The multi-scale features enhanced by PANet improve the boundary accuracy of the mask.

Using a depth camera, we acquire RGB images containing target fruits and canopy branches, along with depth point sets aligned to the pixel coordinate system of the RGB image. The RGB images are fed into the trained Mask-RCNN-FPN3x model to detect and segment fruits and branches, producing segmented mask images for each fruit target, labeled as $fruit_0$、$fruit_1$、 ...、$fruit_i$. The Canny edge detection algorithm is then applied to extract edge points of each fruit in the image coordinate system, forming the contour point set $\{k_1\}\{k_2\}\{k_3\}...\{k_i\}$ for each fruit, The detailed process is illustrated in Figure 4.

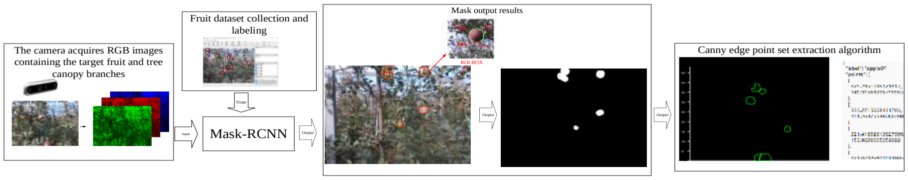
The camera acquires RGB images containing the target fruit and tree canopy branches
Fruit dataset collection and labeling
Train
Input
Mask-RCNN
Output
Mask output results
ROI BOX
Output
Canny edge point set extraction algorithm
Output
"label": "apple0"
"points": [

Figure 4. Pixel-Level Instance Segmentation Visualization Results of Apple Fruits.

3.1.2 Fitting 2D Plane Completion of Fruits Based on Maximum Circumscribed Circle

For the contour point set of apple fruits mentioned earlier, a minimum circumscribed circle can always be found for each fruit contour. The center of this minimum circumscribed circle is denoted as $C_i(x, y)$, and the radius R_i ensures that all inner points p_i of the fruit contour point set $\{k_i\}, i = (1,2,3..., n)$ lie within the boundary and interior of the circumscribed circle centered at $C_i(x, y)$.

Specifically, as illustrated in the figure 5, for each fruit, three convex hull points (outer edge points) are randomly selected from the point set to check whether the current minimum circumscribed circle $C_i(x, y)$contains all points of the fruit contour. If it does not, the radius $R_i$of the current minimum circumscribed circle is increased, and three convex hull points are drawn again. This process is recursively continued and iterated until the following condition is met:

$$R_i = \{minC_i(x, y)_{r_i}\} \tag{9}$$

The constraint condition is satisfied as follows:

$$|p_i - C_i(x, y)|_2 \leq r_i \quad , \quad \forall p_i \in \{k_i\} \tag{10}$$

where $|p_i - C_i(x, y)|$ represents the Euclidean distance from point p_i to the center $C_i(x, y)$, and $\forall p_i \in \{k_i\}$ indicates that all points within the closed contour point set $\{k_i\}, i = (1,2,3..., n)$ satisfy this condition.

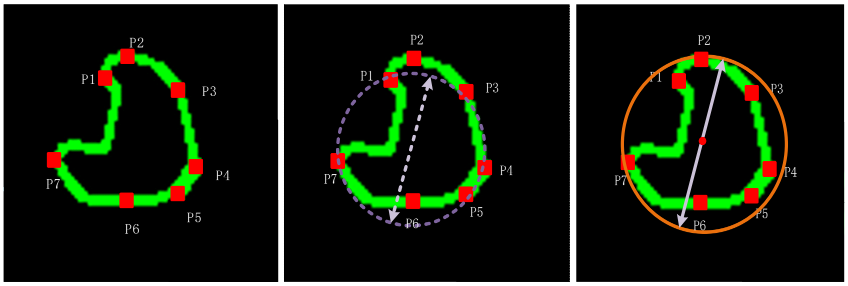

Figure 5. Fitting 2D Plane Completion of Fruits.

3.2 Extraction of Depth Point Set for Occluded Fruits and Prediction of Effective 3D Picking Points

3.2.1 Extraction and Correction of Depth Point Set for Fruits

To extract the depth points of the occluded fruit within the 2D area defined by center $C_i(x, y)$ and radius R_i, the depth point set is given by:

$$\{D_i\} = \left\{ d_{(u,v)} \;\middle|\; \sqrt{(u-x)^2 + (v-y)^2} \leq r \right\} \tag{11}$$

A clustering algorithm divides $\{D_i\}$ into three groups: foreground obstacles, fruit body, and background as C_1, C_2, C_3. The depth values of the fruit body are denoted as C_2.

We calculate the mean depth value μ_i and standard deviation σ_i for the fruit body:

$$\mu_i = \frac{1}{|n_{C_2}|} \sum_{d_{(u,v)} \in C_2} d_{(u,v)} \tag{12}$$

$$\sigma_i = \sqrt{\frac{1}{|n_{C_2}|} \sum_{d_{(u,v)} \in C_2} (d_{(u,v)} - \mu_i)^2} \tag{13}$$

Effective depth values must satisfy:

$$\mu_i - 2\sigma_i \le d \le \mu_i + 2\sigma_i \tag{14}$$

The filtered and corrected effective depth point set of the occluded fruit is denoted as $\{D_i'\}$.

3.2.2 Prediction of Effective 3D Picking Points for Occluded Fruits

For the effective depth point cloud set $\{D_i'\}$ of occluded fruits, we perform geometric sphere feature fitting and 3D center prediction. The error minimization function (E_i) for the center coordinates $O_i(x_i, y_i, z_i)$ of the spheroid fruit is given by:

$$E(x_i, y_i, z_i, r_i) = \sum[(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 - {r_i}^2]^2 \tag{15}$$

where (x, y, z) represents any point on or within the spheroid fruit, and r_i is the estimated 3D radius of the fruit.

By taking partial derivatives of E with respect to (x_i, y_i, z_i, r_i) and setting them to zero, we can fit the correct 3D center position $O_i(x_i, y_i, z_i)$ for each fruit, which serves as the picking localization point.

4. RESULTS AND ANALYSIS

To validate the effectiveness of the proposed algorithm, experiments were conducted using an orchard image dataset that includes occlusion and complex lighting conditions. The dataset was sourced from Siweite Fruit Industry Co., Ltd. in Suqian City, Jiangsu Province, China, and contains fruit images with varying degrees of occlusion (e.g., partial occlusion and tight contact occlusion). The true boundaries and locations of the fruits were annotated and divided into three scenarios based on occlusion ratios of 10%, 20%,30%,40% and 50%. The experiments were carried out on a computer equipped with an RTX 2070 GPU, using the PyTorch deep learning framework and Mask-RCNN as the base instance segmentation model.

4.1 Occluded Fruit Target Recognition and Segmentation Testing Under Dense Canopies

This experiment utilized the Mask-RCNN model for instance segmentation, The model test results for the test set are shown in Figure 6. Aiming to test segmentation accuracy under different occlusion ratios. Edge detection (Canny algorithm) was employed to further refine the fruit boundaries. Average precision (AP) for instance segmentation and F1 Score for edge detection were used as evaluation metrics, with varying occlusion ratios as the variable.

The edge detection accuracy is represented by the F1 Score, calculated as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

where Precision and Recall are defined as:

$Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$.

Here, TP、FP、FN represent the true positives, false positives, and false negatives in edge detection, respectively.

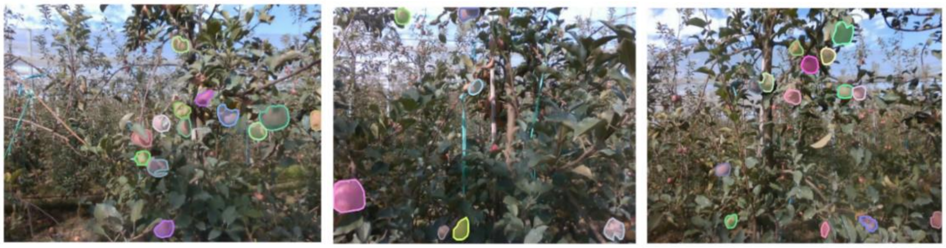

Figure 6. Test set of the model test results.

4.2 Recognition and Segmentation Testing of Occluded Fruits

4.2.1 Evaluation Metrics and Experimental Results

The average precision (AP) was calculated using IoU (Intersection over Union) to measure instance segmentation accuracy. For each segmentation result, IoU is defined as:

$$IoU = \frac{|M_{pred} \cap M_{true}|}{|M_{pred} \cup M_{true}|} \tag{17}$$

whereM_{pred} is the predicted mask area, andM_{true} is the actual mask area.

The results for mAP at different IoU thresholds, including the more lenient mAP@0.5 and the stricter mAP@0.5:0.95, as well as edge detection F1 Score, are summarized in the following table:

Table 2: Object Detection Model Test Results

Occlusion Ratio	**Instance Segmentation AP (%)**	**Edge Detection F1 Score**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
10%	96.8	0.91	94.5	82.3
20%	95.4	0.89	91.7	79.6
30%	92.4	0.87	88.1	74.6
40%	90.2	0.84	86.0	71.5
50%	88.3	0.80	83.7	67.2

The model testing results based on the test set are shown in Figure 6.

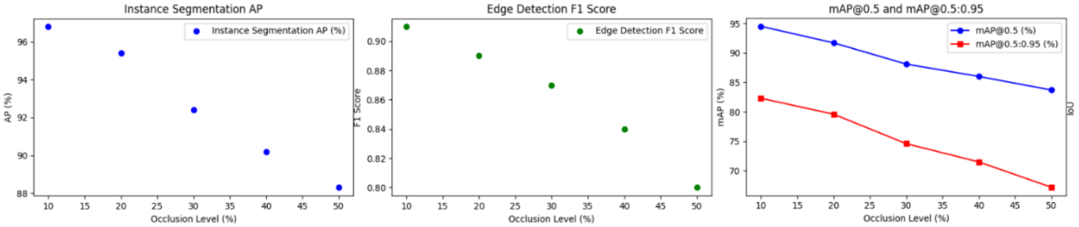

Figure 7. Object Detection Model Test Results

4.2.2 2D Completion and 3D Restoration of Occluded Fruit Targets Under Dense Canopies

To assess the accuracy of completing the 2D planar contours of occluded fruits, we used the point set of the segmented fruit contours and applied the maximum circumscribed circle fitting method to generate the completed 2D contour. The 2D completion accuracy (2D Completion Accuracy) was calculated by comparing the completed circular area with the actual occluded area; a higher IoU value indicates better completion.

The 2D completion accuracy is calculated as:

$$IoU_{2D} = \frac{|R_{pred} \cap R_{true}|}{|R_{pred} \cup R_{true}|} \tag{18}$$

where R_{pred} is the generated completed circular area and R_{true} is the actual occluded area.

To evaluate the accuracy of fitting the geometric sphere characteristics and estimating the 3D center of the occluded fruits, we extracted the depth point cloud of the occluded fruits from depth images. The 3D restoration accuracy was assessed by calculating the Euclidean distance between the estimated 3D center position and the true center position, referred to as the 3D restoration error $\varDelta Error_{3D}$:

$$\varDelta Error_{3D} = \sqrt{(x_c - x_{true})^2 + (y_c - y_{true})^2 + (z_c - z_{true})^2} \tag{19}$$

where (x_c, y_c, z_c) denotes the computed estimated 3D center position, and $(x_{true}, y_{true}, z_{true})$ is the true center position.

The results of the 2D completion accuracy and 3D reconstruction error are summarized in the following table:

Table 3: 2D Completion Accuracy and 3D Reconstruction Error Table:

Occlusion Ratio	**2D Completion Accuracy (IoU)**	**3D Restoration Error (mm)**
10%	0.93	3.1
20%	0.91	4.0
30%	0.89	5.4
40%	0.85	6.5
50%	0.82	7.6

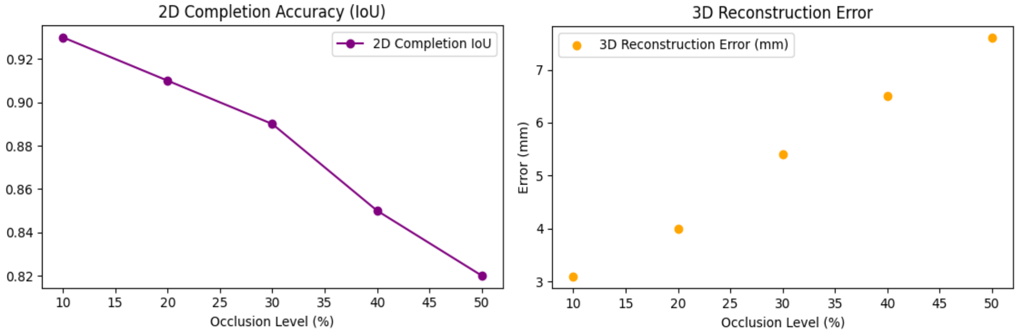

Figure 8. 2D Completion Accuracy and 3D Reconstruction Error results

5. CONCLUSION

In semi-structured environments like orchards with dense canopies, accurately identifying and locating target fruits while sensing surrounding obstacles is a pressing challenge in harvesting operations. This study presents an efficient method for recognizing, segmenting, and locating occluded fruit targets specifically for dense canopy environments. By leveraging the Mask R-CNN model combined with multi-scale feature fusion using FPN and PANet, along with improved 2D and 3D completion algorithms, precise segmentation and localization of fruits under occlusion conditions have been achieved.

Experimental results show that at an occlusion rate of 10%, the model performs excellently across all evaluation metrics, with an average precision (AP) of 96.8% and an edge detection F1 score of 0.91. Even with an increased occlusion rate of 50%, the model maintains high segmentation accuracy, achieving an AP of 88.3% and an edge F1 score of 0.80. Additionally, the implementation of 2D completion and 3D reconstruction algorithms enables the contour completion of occluded fruits and the 3D localization of picking points, with a 3D reconstruction error of only 3.1 mm at 10% occlusion and 7.6 mm at 50% occlusion. These results demonstrate that the proposed method exhibits good robustness and practicality in complex environments, facilitating successful fruit harvesting tasks.

This research provides an effective technical framework for fruit recognition and harvesting robot systems in dense canopy environments. Future work will focus on further optimizing the algorithm's computational efficiency to meet real-time demands and exploring multi-sensor fusion techniques to enhance detection accuracy and stability in highly occluded scenarios.

REFERENCES

[1] ZHAO Chunjiang, FAN Beibei, LI Jin, FENG Qingchun. Agricultural Robots: Technology Progress, Challenges and Trends. Smart Agriculture, 2023, 5(4): 1-15.

[2] Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. Remote Sens. 2021, 13, 1619.

[3] Fangfang Gao, Longsheng Fu, Xin Zhang, Yaqoob Majeed, Rui Li, Manoj Karkee, and Qin Zhang. "Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN." *Computers and Electronics in Agriculture*, vol. 176, 2020, p. 105634. ISSN 0168-1699.

[4] Lv Jidong, Zhao De-An, Ji Wei, and Ding Shihong. "Recognition of apple fruit in natural environment." *Optik*, vol. 127, no. 3, 2016, pp. 1354-1362.

[5] Peng Y, Zhao S, Liu J. Segmentation of overlapping grape clusters based on the depth region growing method. Electronics, 2021, 10(22): 2813.

[6] Xu, Zhujie, Jizhan Liu, Jie Wang, Lianjiang Cai, Yucheng Jin, Shengyi Zhao, and Binbin Xie. "Realtime Picking Point Decision Algorithm of Trellis Grape for High-Speed Robotic Cut-and-Catch Harvesting." *Agronomy* 13, no. 6 (2023): 1618.

[7] Li, Li, Zhi He, Kai Li, Xinting Ding, Hao Li, Weixin Gong, and Yongjie Cui. "Object detection and spatial positioning of kiwifruits in a wide-field complex environment." *Computers and Electronics in Agriculture*, vol. 223, 2024, p. 109102.

[8] Changqing Gao, Hanhui Jiang, Xiaojuan Liu, Haihong Li, Zhenchao Wu, Xiaoming Sun, Leilei He, Wulan Mao, Yaqoob Majeed, Rui Li, and Longsheng Fu. "Improved binocular localization of kiwifruit in orchard based on fruit and calyx detection using YOLOv5x for robotic picking." *Computers and Electronics in Agriculture*, vol. 217, 2024, p. 108621.

[9] Longsheng Fu, Fangfang Gao, Jingzhu Wu, Rui Li, Manoj Karkee, and Qin Zhang. "Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review." *Computers and Electronics in Agriculture*, vol. 177, 2020, p. 105687.

[10] Chu P, Li Z, Lammers K, et al. Deep learning-based apple detection using a suppression mask R-CNN. Pattern Recognition Letters, 2021, 147: 206-211.

[11] Yang Yu, Kailiang Zhang, Li Yang, and Dongxing Zhang. "Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN." *Computers and Electronics in Agriculture*, vol. 163, 2019, p. 104846.

[12] LIU, Jizhan, ZHU, Xinxin, and YUAN, Yan. "Depth-sphere Transversal Method for on-branch Citrus Fruit Recognition." *Transactions of the Chinese Society for Agricultural Machinery*, vol. 48, no. 10, 2017, pp. 32-39.