

The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan

PRECISE STRAWBERRY STEM LOCALIZATION VIA TWO-STAGE 3D DEEP LEARNING

Rui-Yi Xu, Chia-Jung Chang, Chung-Liang Chang*

Department of Biomechatronics Engineering,
National Pingtung University of Science and Technology,
Pingtung county, TAIWAN.

*Corresponding Author: chungliang@mail.npust.edu.tw

ABSTRACT

Harvesting delicate fruits, such as strawberries, at their optimal ripening stage inherently presents significant challenges, given its labor-intensive, time-consuming nature and high susceptibility to mechanical damage. The solution lies in developing intelligent robotic harvesting systems that can accurately segment fruits, determine optimal picking locations, and perform delicate maneuvers without compromising fruit integrity. This study presents a novel 3D vision-based methodology for autonomous cutting point detection in fragile fruits. 3D imaging method is used to obtain the appearance features of the fruit. These features are subsequently processed by a custom-designed deep learning architecture, integrating Top-down and Bottom-up processing schemes, to accurately segment ripe fruit instances, precisely define their contours, and ultimately pinpoint the optimal cutting location. With strawberries, the model achieved an average recognition accuracy of 67% (mAP 0.5-0.95). Furthermore, the detected picking point positioning showed a mean absolute error of only 0.18 cm.

Keywords: 3D vision, deep Learning, strawberry harvesting, cutting point detection, robotic harvesting.

INTRODUCTION

Strawberry harvesting is labor-intensive, time-consuming, and prone to fruit damage due to the crop's delicate texture and irregular shape. The growing shortage of agricultural labor underscores the need for intelligent robotic harvesting systems. A key challenge for such systems lies in the real-time identification of ripe fruit and accurate localization of the optimal picking point (Yiong et al., 2019; Chang and Huang, 2024). This study introduces a multi-task detection and segmentation framework for strawberry plants using the state-of-the-art YOLO11 model, representing its first application in this domain. The model reliably identifies strawberries at different maturity stages (unripe and ripe) and detects surface diseases.

MATERIALS AND METHODS

Our robotic system adopts an eye-in-hand configuration, where the visual sensor is mounted

directly on the end-effector of the robotic arm to enable high-precision, close-range perception. A six-axis robotic arm with high accuracy and repeatability is employed to perform stable and flexible picking motions. An Intel RealSense D405 depth camera serves as the eye-in-hand vision system. Operating at 720p resolution and 30 frames per second (FPS), it provides clear real-time RGB and depth data, making it well suited for close-range harvesting tasks.

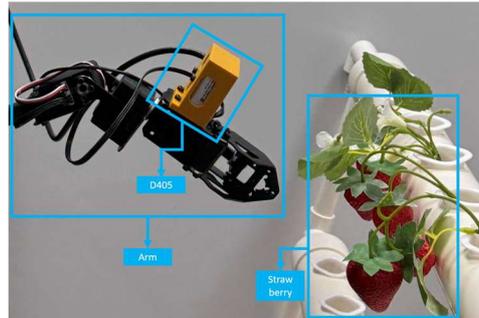


Fig.1 Experimental platform configuration.

This core of this study is to convert visual information into executable 3D robot coordinates. The process combines a top-down and bottom-up approach, integrating deep learning with a custom post-processing algorithm. First, the top-down stage uses a pre-trained YOLO11 model on RGB images from a D405 camera to perform object detection and generate instance segmentation masks. This accurately identifies ripe, unripe, and diseased strawberries, providing essential 2D data. Next, the bottom-up stage develops an algorithm to locate the precise harvesting point. By utilizing the depth map and pixel-level masks, the algorithm identifies the stem-to-fruit connection point with an accuracy of up to 0.18 cm, ensuring damage-free picking.

RESULTS & DISCUSSION

The proposed method was evaluated on an independent test set, with key performance metrics shown in Table 1. For the task of ripe strawberry harvesting, our model achieved a mean precision (P) of up to 94.7% and a mean recall (R) of 95% in both object detection and instance segmentation. More importantly, by combining depth information with post-processing, our picking point localization algorithm achieved an exceptionally high level of accuracy, with a mean absolute error (MAE) of only 0.18 cm. This demonstrates that the robotic gripper can align with the target with extremely high precision.

Table 1 Model performance evaluation on different classes.

Class	Instances	Box mAP@0.5-0.95	Mask mAP@0.5-0.95
all	3819	0.759	0.584
unripe	1452	0.659	0.461
ripe	1861	0.831	0.703
surface_disease	506	0.787	0.587

To provide a more intuitive demonstration of the model's performance, several real-world detection and localization examples are presented in Fig. 2. As shown, the model accurately labels ripe strawberries with bounding boxes and instance segmentation masks. These results

visually confirm the reliability of the multi-task perception framework and the eye-in-hand scheme for practical applications.



Fig. 2 Model detection and area estimation.

CONCLUSIONS

This study presents a high-performance visual perception system that integrates an eye-in-hand vision scheme with the YOLO11 deep learning model. The core contribution is a customized post-processing algorithm that fuses high-precision instance segmentation results with depth information. This approach achieves ultra-precise strawberry stem localization with a mean absolute error of 0.18 cm. Future work will focus on evaluating the system through real strawberry harvesting experiments and analyzing its performance.

ACKNOWLEDGEMENTS

This research was funded by the Agriculture and Food Agency, Ministry of Agriculture, Taiwan, under grant number 114AS-4.5.1-FD-01(3).

REFERENCES

- Chang, C. L., and C. C. Huang. 2024. Design and implementation of an AI-based robotic arm for strawberry harvesting. *Agriculture* 14(11): 2057. <https://doi.org/10.3390/agriculture14112057>
- Yiong, Y., C. Peng, L. Grimstad, P. J. From, and V. Isler. 2019. Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. *Computers and Electronics in Agriculture* 157: 392-402. <https://doi.org/10.1016/j.compag.2019.01.009>.