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

NULL DATASET-BASED DETECTION ENHANCES ROBOTIC VISION IN GREENHOUSE CHERRY TOMATO HARVESTING

Verianti Liana[1], Sakir Kanmis[2], Shang-Wei Hsu[2], Chang Chiu[2], Ping-Lang Yen[1,2*]
[1]Global Agriculture Technology and Genomic Science Program, International College, National Taiwan University, Taiwan.
[2] Department of Biomechatronics Engineering, National Taiwan University, Taiwan.

*Corresponding Author: plyen@ntu.edu.tw

ABSTRACT

Cluttered cherry tomato greenhouse environments with visually similar distractors often trigger False Positives (FPs) in robotic vision, misguiding the robot's motion and reducing harvesting success. We introduce a null-dataset strategy that integrates unannotated distractor images into YOLOv8l training, with their proportion tuned through loop refinement to suppress FPs while preserving precision. Optimal null proportions were identified as 12.3% for tomato detection and 8.3% for pedicel detection, with improvements validated by one-tailed Z tests at α = 0.01 (p < 0.00001). The refined models were deployed on a UGV in real-time greenhouse trials at 30 FPS. Across 20 trials, all targeted cherry tomatoes were detected precisely, and 90% of pedicels with calyx and cutting point were localized without FPs. These results demonstrate that null dataset integration provides a practical strategy to enhance precision and reliability in greenhouse harvesting robots without increasing model complexity or latency.

Keywords: Robotic vision, Null dataset, False positive suppression, Precision agriculture, Object detection

INTRODUCTION

Robotic harvesting in greenhouses requires vision systems that remain precise in cluttered environments. In Taiwan's cherry tomato production, red grafting clips, about 15 clippers per meter, mimic ripe fruit, while leaves, calyxes, stems, and peduncles share visual cues with the pedicel and its cutting point. Since market standards demand calyx retention as a signal of freshness, precise detection of both calyx and cutting point is essential. Yet direct identification of these fine structures during initial scanning remains highly challenging. To address this, our pipeline first detects fully ripe tomatoes, then localizes the pedicel and cutting point for harvesting. Errors at the detection stage cascade into incorrect cutting and reduce success. Object detection, therefore, forms the pipeline's foundation, with YOLOv8 widely adopted for real-time use but still prone to false positives (FPs) in cluttered scenes (Wu et al., 2021). Instead of adding architectural complexity, which risks latency and unstable FPS on embedded robots (Li et al., 2025), we retain YOLOv8l and suppress FPs using a loop-refined null dataset of unannotated distractor images. The effect is assessed with one-tailed Z tests at α = 0.01 and validated in real-time robotic trials.

This study evaluates, in real-time greenhouse harvesting trials, whether integrating a loop-refined null dataset into YOLOv8l suppresses FPs and enhances precision, with statistical significance assessed using one-tailed Z tests at α = 0.01.

MATERIALS AND METHODS

Our Unmanned Ground Vehicle (UGV) harvesting pipeline first detects fully ripe cherry tomatoes (redness ≥ 90%) at about 35 to 40 cm, then moves to ~20 cm for confirmation and viewpoint search, and finally performs pedicel detection to localize the cutting point for calyx-retaining harvest (Fig. 1). Detection occurs at three stages (Fig.1) and directly affects harvesting success.

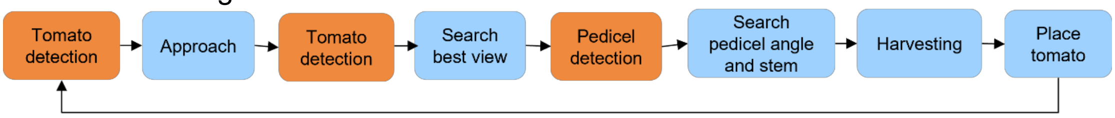

Fig 1. Overview of Cherry Tomato Harvesting Pipelines

We trained a YOLOv8l tomato detection model on 4,767 images with 12.3% null dataset (300 epochs, batch size= 64, SGD optimizer, lr_0= 0.0005, weight decay=0.006) and a YOLOv8l-Pose pedicel detection model on 2,610 images with 8.3% null data (300 epochs, batch size= 64, AdamW optimizer, lr_0= 0.005). The pedicel dataset was annotated with key points (calyx and its cutting point) across varied viewpoints, occlusions, clusters, and distractors. In both tasks, null data proportions were refined iteratively: frequently misclassified items were added as unannotated nulls, and a new proportion was accepted only when a one-tailed two-proportion Z test (α = 0.01) confirmed significant FP reduction and precision gain. Evaluation was conducted on 570 testing images across 19 scenarios for tomato detection and 360 images across more than 10 scenarios for pedicel detection. The refinement objective prioritized FP suppression while maximizing precision:

$$\theta^* = \arg\max_{\theta} [\, \text{Precision}(\theta) - \lambda \cdot \text{FDR}(\theta)] \quad (1)$$

where θ is the null data proportion in a given refinement round, Precision(θ) is the fraction of correct detections, False Discovery Rate (FDR(θ)) is FP / (True Positive + FP), and λ > 1 penalizes FPs due to costly robotic mis-movements. Both models with integrated null datasets were executed in real time on a UGV equipped with an NVIDIA AGX Orin during 20 harvesting trials along a 28-meter plant row in a Taoyuan greenhouse. In real-time deployment, the first few frames that satisfy the fully ripe cherry tomato criterion are forwarded to trigger the approach and pedicel cutting, so the precise detection is critical. These same forwarded frames are used to compute real-time precision (with an allowable error distance of 5 mm) and FDR.

RESULTS & DISCUSSION

Using the loop-refinement strategy, the optimal null dataset proportions were 12.3% for tomato detection and 8.3% for pedicel detection. With null integration, tomato precision increased from 0.8958 to 0.9622 (Z = +9.59) and FDR decreased from 10.42% to 3.78% (Z = –9.60). For pedicel detection, precision rose from 0.8430 to 0.9718 (Z = +10.47) while FDR dropped from 15.70% to 2.82% (Z = –10.48). All Z tests yielded $p < 0.00001$, confirming that null data significantly improves precision and suppresses FPs. This test demonstrates that the model with null dataset effectively avoids detecting distractors, highlighted by the black-circled objects in Fig. 2. When no pedicel is present in the frame, the model with a null dataset successfully produces no detections, whereas the model without null dataset yields incorrect and random detections. The refined models were deployed on a UGV for real-time greenhouse trials at 30 FPS (Fig. 3). Within 10–20 frames, fully ripe cherry tomatoes and most pedicels were detected. Across 20 trials, all target tomatoes were detected despite clutter from red clippers, and 90% of pedicels with calyx and cutting point were localized without FPs. The remaining 10% of pedicels were undetected but produced no FPs (0 FDR), demonstrating that the model detects only when highly confident, ensuring precision in real-time harvesting.

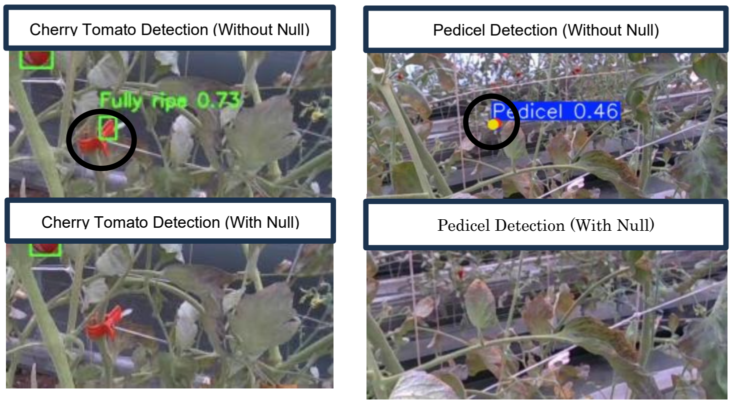

Fig. 2. Comparison of fully ripe cherry tomato and pedicel detection with and without null dataset integration. Black circles highlight FPs caused by red clipper (cherry tomato distractor) and stem (pedicel distractor).

Fig. 3. Real-time detection in greenhouse trials: pedicel (left, red point = pedicel cutting point, green point = calyx) and fully ripe cherry tomato (right).

ACKNOWLEDGEMENTS

The paper is financially supported by the Ministry of Agriculture, Taiwan under grant 114AS-16.1.1-AS-04.

CONCLUSIONS

Integrating loop-refined null datasets of 12.3% for cherry tomato detection and 8.3% for pedicel detection significantly reduced FPs and improved precision. Real-time trials confirmed 100% tomato and 90% pedicel detection precision with no FPs and within a 5 mm allowable error distance, showing that null data integration is an effective and practical strategy to strengthen greenhouse harvesting robots without adding model complexity.

REFERENCES

Li, S., Z. Chen, J. Xie, H. Zhang, and J. Guo. 2025. PD-YOLO: a novel weed detection method based on multi-scale feature fusion. Frontiers in Plant Science. 16: 1-20.

Wu, Z., Y. Chen, B. Zhao, X. Kang, and Y. Ding. 2021. Review of weed detection methods based on computer vision. Sensors. 21(11): 1-23.