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

SYNTHETIC DATA-DRIVEN VALIDATION OF MULTI-STAGE FRUIT DETECTION SYSTEMS IN CONTROLLED VIRTUAL ENVIRONMENTS

Yan-An Chen, Ssu-Chi Chen, Shih-Fang Chen*

Department of Biomechatronics Engineering, National Taiwan University, Taipei, Taiwan

*Corresponding Author: sfchen@ntu.edu.tw

Abstract

Accurate fruit counting across development stage is critical for tomato breeding decisions. Yet, the ground truth validation in real field remains challenging where partially occluded fruits cannot be reliably counted manually due to complex environmental factors. To address this need, this study presents a photorealistic simulation approach that complements real field data collection. A virtual environment enables controlled evaluation across three distinct fruit growth stages: green stage fruit, turning-pink stage fruit, and harvest stage fruit. This system integrates three components: a Plant Recognize System for plant region selection where the tomato plant appears, a YOLOv11-based detection model for multi-stage fruit identification, and a Fruit Counting Algorithm for temporal counting. Over 2,000 tomato plant images were annotated to train the detection model. The simulation environment provides the complete ground truth of fruit numbers per plant, enabling accurate performance assessment difficult to achieve in field environments. This controlled approach allows customized evaluation across different lighting conditions, plant densities, and growth stages. Field validation demonstrated that YOLOv11 detection model achieved an average precision of 86.7% and a recall of 80%. Furthermore, the tracking algorithms yielded mean absolute errors (MAE) of 4.42 in green stage fruit, 0.95 in turning-pink fruit, and 1.06 in harvest stage fruit. Performance variations across different conditions were successfully identified. The simulation-based approach showed a positive step towards the use of virtual environment for agricultural computer vision development, enabling robust model training and validation without extensive field data collection requirements. The framework supports scalable phenotypic monitoring system development for precision breeding applications.

Keywords: Virtual environment, Fruit counting, Photorealistic simulation, Synthetic data

INTRODUCTION

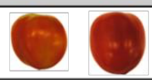
Multi-stage tomato fruit counting is essential for breeding decisions, but field-based ground truth validation presents significant challenges due to occlusion and environmental complexity. Traditional development cycles require substantial time for data collection across the brief maturity stage windows (3-6 days each), while field conditions vary significantly in illumination, occlusion, and canopy density. To address this, this study proposes a simulation-in-the-loop framework using a controllable 3D virtual farm built with open-source simulation tools. The approach shifts from conventional field-only development to a faster, controllable simulation-guided workflow that provides standardized validation. Using this virtual testbed, we developed a complete multi-stage fruit counting system delivering robust per-plant counting while maintaining practical inference speed.

MATERIALS AND METHODS

DATASET COLLECTION

Approximately 3,000 tomato images were collected at the World Vegetable Center (Tainan, Taiwan). Fruits were categorized into three ripening stages: green (early development), turning-pink (transition phase with partial yellow-pink coloration), and harvest (mature red surface) (Table 1). Data was split 8:1:1 for training, validation, and testing.

Table 1 Tomato fruit ripening stages and instance counts in datasets.

Fruit stage / Object	Green stage fruit	Turning-Pink stage fruit	Harvest stage fruit
	Color change: 0%	Color change: 1%~60%	Color change: 60%~100%
Image			
Instance	13073	1983	3264

VIRTUAL ENVIRONMENT DEVELOPMENT

Tomato plant models were constructed in Blender (version 4.4) using three components: main vine, developmental-stage leaves, and maturity-stage fruits. A parametric workflow employs a single guide curve defining the vine with procedurally attached leaf and fruit instances. Model parameters control fruit counts per stage, leaf density, and morphology, enabling rapid plant configuration. Virtual fields were assembled and simulated in NVIDIA Isaac Sim. Solar parameters (elevation, azimuth, intensity) varied to simulate different lighting conditions. A simulated agricultural vehicle with onboard cameras navigated predefined field trajectories, providing repeatable datasets for algorithm evaluation (Fig. 1).

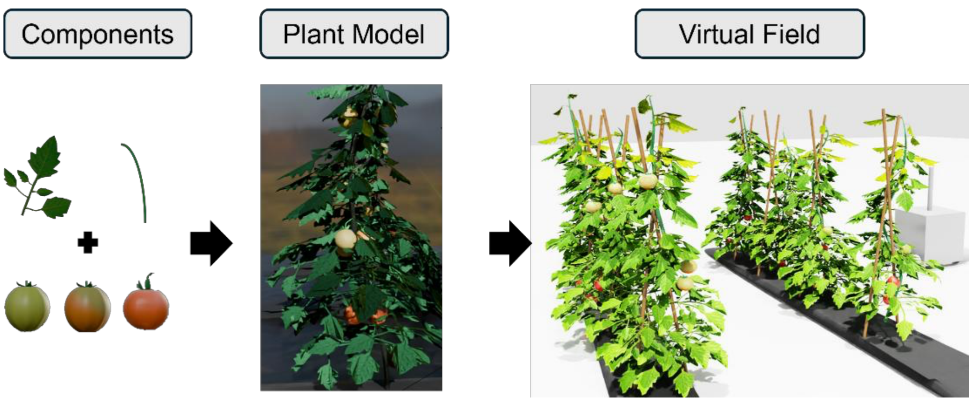


Fig. 1 Stage-aware plant model and simulated farm environment.

SYSTEM ARCHITECTURE

The counting system comprises two modules: (i) plant-region recognition using monocular depth, and (ii) multi-stage fruit detection, tracking, and counting within plant regions (Fig. 2).

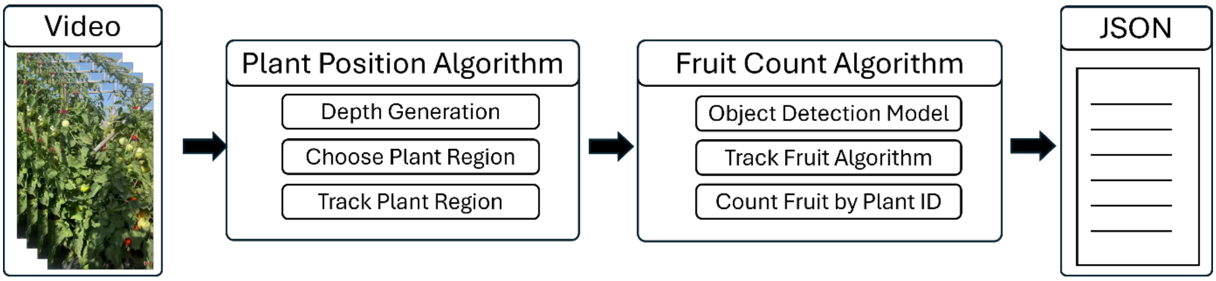


Fig. 2 Architecture of the tomato fruit counting system.

PLANT POSITION DETECTION

Side-view imagery exhibits consistent gaps between adjacent tomato plants. Monocular depth estimation using Depth Anything v2 (Yang et al., 2024) generates per-frame depth maps $D \in R^{H \times W}$. Vertical projection produces a gap-revealing signature:

$$d(x) = \frac{1}{H} \sum_{y=1}^{H} D(y, x) \quad (1)$$

Gaussian filter smooths the depth profile *d(x)* while preserving broad depth structure. Local extrema indicate inter-plant gap transitions; contiguous spans between valleys define plant regions. Lightweight tracking maintains temporal continuity and suppresses spurious detections.

FRUIT DETECTION AND CONTING

YOLOv11(Khanam et al., 2024) classifies fruits across three maturity stages. Single-frame detections are unreliable in field conditions due to leaf and fruit occlusion; therefore, multi-object tracking associates detections across frames to recover occluded instances and prevent double counting. Detections within tracked plant regions are attributed to corresponding plants. Stage-wise counts accumulate per plant using persistent track IDs.

RESULTS & DISCUSSION

Virtual environment testing demonstrated robust performance across viewpoints and illumination conditions in the 3D virtual farm. Field validation showed the plant-region detector achieved a mean absolute error (MAE) of 6.1 plants per video. YOLOv11 attained an average precision (AP) of 86.7% and a recall of 80% for fruit detection on the field dataset, respectively. Fruit counting yielded MAEs of 4.42 (green stage), 0.95 (turning-pink), and 1.06 (harvest stage). High green-stage errors primarily resulted from class imbalance between green and turning-pink categories (Fig. 3). The brief turning-pink duration (3 days) creates under-representation in datasets, causing misclassification of turning-pink as green fruits. Tracking performance degraded when temporal continuity was disrupted by abrupt camera motion or severe occlusion, affecting fruit-to-plant attribution accuracy. Overall, the virtual farm approach accelerates algorithm design and validation effectively.

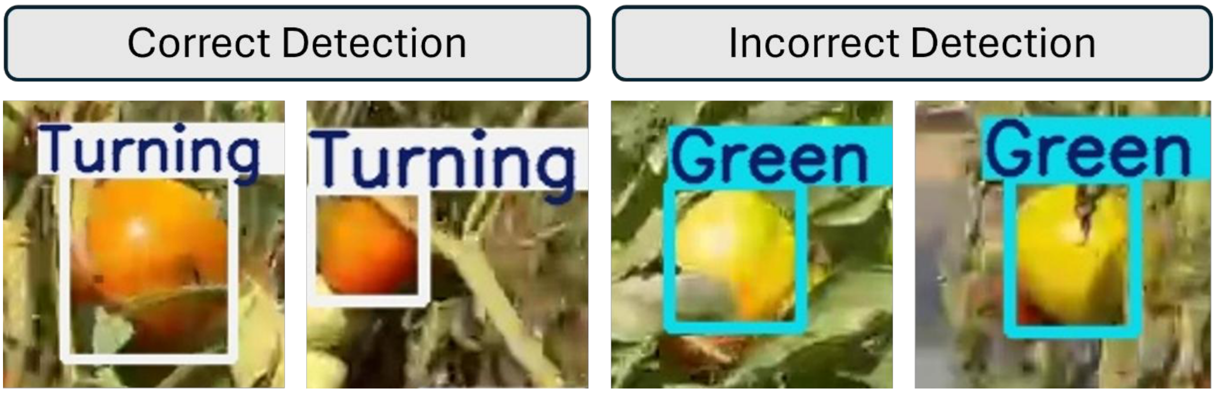


Fig. 3 Demonstration of YOLOv11 detection in tuning-pink stage fruit.

CONCLUSIONS

This study demonstrates that 3D virtual tomato farms constructed with computer graphics enable controlled, repeatable experiments across viewing conditions and lighting scenarios. The simulation addresses constraints from short crop growth windows and reduces field data collection requirements for algorithm development. Algorithms validated in virtual environments exhibit stronger robustness, facilitating rapid generalization to new cultivars and field configurations. These results indicate a practical pathway for faster, more reliable agricultural algorithm deployment.

ACKNOWLEDGMENTS

This study acknowledges the assistance provided by the staffs at the World Vegetable Center (Tainan City, Taiwan). The research was supported by the National Science and Technology Council, Executive Yuan, Taiwan (Project No. 113-2313-B-002-030-MY3).

REFERENCES

Mirbod, O., Choi, D., & Schueller, J. K. (2025). From Simulation to Field Validation: A Digital Twin-Driven Sim2real Transfer Approach for Strawberry Fruit Detection and Sizing. *AgriEngineering*,*7*(3), 81.

Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., & Zhao, H. (2024). Depth anything v2. *Advances in Neural Information Processing Systems*, *37*, 21875-21911.

Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.

Aharon, N., Orfaig, R., & Bobrovsky, B. Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*.