APPLICATION OF DEEP LEARNING FOR SYMPTOM DETECTION AND LOCALIZATION IN PHALAENOPSIS PLANTLETS

Zhi-Yuan Chen[1], Hong-Sheng Syu[2], Yu-Sheng Zeng[1], Ying-Jen Huang[1*]

[1] Department of Biomechatronics Engineering, National Chiayi University, Chiayi, Taiwan
[2] Department of Computer Science and Information Engineering, National Chiayi University, Chiayi, Taiwan
*Corresponding Author: jen@mail.ncyu.edu.tw

ABSTRACT

Phalaenopsis plantlets in dense greenhouses are vulnerable to diseases like soft rot, which spreads rapidly. This study compares YOLOv11 with enhanced architectures (FasterNet, MambaVision) for symptom detection and localization. Single- and multi-model strategies were evaluated for disease recognition, plant segmentation, and keypoint localization, enabling robotic removal and efficient automated disease management.

Keywords: Phalaenopsis plantlet, disease symptom detection, plant location, YOLO, FasterNet, MambaVision

INTRODUCTION

Phalaenopsis plantlets grown in dense greenhouses are vulnerable to fast-spreading diseases like soft rot, making manual inspection inefficient.

This study presents an automated system that detects symptoms, associates them with individual plants, and identifies grasping points, enabling robotic removal and improving greenhouse management.

MATERIALS AND METHODS

2.1 Image Collection

A gantry-style imaging platform was designed based on greenhouse bench size to capture top-view images of Phalaenopsis plantlets. Additional images were collected manually using a handheld camera. All images were categorized as healthy or infected with four common diseases—soft rot, anthracnose, yellow leaf, and blight—to construct a symptom dataset.

2.2 Model Architectures

This study adopted YOLOv11 (Khanam & Hussain, 2024)as the baseline detector and replaced its backbone with two alternatives: **FasterNet** (Chen et al., 2023) (partial convolution for efficient inference) and **MambaVision** (Hatamizadeh & Kautz, 2025) (a hybrid of CNN, state-space modeling, and attention for long-range dependency). Models were trained under the same conditions (500 epochs, batch size 12, image size 640).

2.3 Tasks and Annotations

Four detection tasks were designed:

1. **Plantlet segmentation** – instance segmentation of plantlet boundaries including leaves, potting media, and pots.
2. **Plantlet pose detection** – bounding boxes plus five keypoints (center, top/bottom medium, left/right leaf tips).
3. **Symptom region detection** – bounding boxes of four symptom types.

4. **Symptom–plantlet pose detection** – bounding boxes plus three keypoints (center, top/bottom medium) with disease class labels.

2.4 Detection Strategies

Two strategies were compared:

- **Multi-model matching** – outputs of symptom detection, segmentation, and pose models were combined by overlap analysis to assign symptoms to plantlets.
- **Single-model detection** – symptom-specific plantlet pose detection directly provided disease type, plant location, and grasping points.

RESULTS & DISCUSSION

3.1 Model Performance

All models achieved over **82% mAP** across tasks, while plantlet segmentation and keypoint detection exceeded **96% mAP**. FasterNet slightly improved accuracy without reducing inference time, whereas MambaVision boosted keypoint accuracy (~+10%) but doubled inference time

3.2 Classification Accuracy

On 200 external test images (307 diseased plantlets), **single-model detection** consistently outperformed **multi-model matching**. Single-model achieved over **94% detection and classification accuracy**, while multi-model suffered from error accumulation across sub-models, leading to lower reliability

3.3 Localization Accuracy

For localization, single-model detection achieved pixel errors of about 5–8 px, compared with 25–35 px for multi-model matching. This demonstrates that direct detection is more precise for robotic grasping and removal

CONCLUSIONS

YOLOv11 was compared with FasterNet and MambaVision for symptom detection and localization in Phalaenopsis plantlets. Single-model detection outperformed multi-model approaches in accuracy and efficiency. FasterNet balanced speed and accuracy, while MambaVision improved keypoint precision but was slower. Overall, YOLOv11 with FasterNet offers a practical solution for automated disease monitoring and robotic removal, enhancing greenhouse management efficiency.

ACKNOWLEDGEMENTS

REFERENCES

Chen, J., Kao, S. H., He, H., Zhuo, W., Wen, S., Lee, C. H., & Chan, S. H. G. (2023). Run, don't walk: Chasing higher FLOPS for faster neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12021–12031). IEEE.

Hatamizadeh, A., & Kautz, J. (2025). MambaVision: A hybrid mamba-transformer vision backbone. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 25261–25270). IEEE.

Khanam, R., & Hussain, M. (2024). YOLOv11: An overview of the key architectural enhancements. arXiv. https://arxiv.org/abs/2410.17725