Online ISSN: 3104-9338
Print ISSN: 3104-932X

Edge-Cloud Collaborative Multi-Target Detection and Growth Prediction for Greenhouse Strawberries Based on YOLOv8 and Multimodal Fusion

Yuhan Zhang, Yixi Li, Xu Wang*

Chengdu College of University of Electronic Science and Technology of China

Received: June 28, 2026

Revised: June 29, 2026

Accepted: June 30, 2026

Published online: July 5, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 2, No. 8 (August 2026)

* Corresponding Author: Xu Wang (34106831@qq.com)

Abstract. To address the problems of low manual inspection efficiency and delayed disease detection in greenhouse strawberry cultivation, this paper constructed a multi-target visual detection model based on the YOLOv8 deep learning algorithm, which simultaneously achieved accurate identification of four types of targets: strawberry ripeness, diseases, pests and growth stages. The research dataset was built by combining images continuously collected from field greenhouses and public agricultural datasets, with a total of 17,000 images. Controlled experiments were conducted after image preprocessing, manual annotation and model training. The lightweight model was deployed on the K230 embedded chip to realize local edge inference, and integrated with environmental sensing data through the Vision-Environment-Terminal Network (VET-Net) multimodal fusion architecture to achieve 7-day crop growth trend prediction. Experimental results showed that the identification accuracy of all four visual targets exceeded 94%, and the end-to-cloud data transmission delay was controlled within 2 seconds. The whole scheme was deployed on an autonomous inspection robot, and farmers can obtain real-time detection results through a WeChat Mini Program. The system effectively reduced manual inspection workload and enabled early disease risk prevention, which possesses good practical application value for reducing strawberry planting costs and increasing yield per unit area.

Keywords: *Greenhouse Strawberry; YOLOv8; Machine Vision; Multi-target Detection; Multimodal Fusion; Smart Agriculture*

1. Introduction

Building a strong agricultural country is the foundation of building a great modern socialist country, and facility agriculture is the development direction of China's modern agriculture. China has the largest facility agriculture area in the world, and agricultural robots and intelligent sensing technologies have become the main driving forces for the transformation and upgrading of facility agriculture [1]. In the early stage, the project team conducted investigations on the management methods of strawberry greenhouses, and found that manual inspection is currently the dominant mode in greenhouses, which suffers from high labor costs, low efficiency, and high risk of missed detection, thus causing the optimal prevention and control window to be missed.

In recent years, the rapid development of machine vision and other technologies has opened up a mature path for intelligent agricultural detection. Xue X et al. conducted a systematic study on deep learning-based detection of fine-grained plant diseases, confirming that deep learning has replaced traditional image processing as the mainstream method for agricultural disease identification [2]. Scholars worldwide have conducted extensive research in different application scenarios. For example, Peng W et al. verified the accurate identification of small pests in hilly crops based on improved YOLOv7 [3]; Chen D et al. realized automatic detection of tea pests and diseases based on YOLOv5s [4]; Xu Y et al. completed the lightweight improvement of YOLOv8 and real-time greenhouse detection experiments for tomato leaf diseases [5]. Mao S et al. systematically sorted out the evolution route and technical pros and cons of YOLO series object detection algorithms [6]; Wan Y et al. summarized the general technical framework of YOLO-based visual identification of crop pests and diseases [7]. Zhou Y et al. realized accurate detection of crop disease leaves based on a multimodal feature alignment architecture, confirming that multi-source fusion of vision and environmental data can improve prediction accuracy [8].

However, existing studies mostly focus on the optimization of single visual disease detection algorithms or the accuracy improvement of multimodal feature fusion. For strawberry greenhouse scenarios, there are still three notable limitations: most detection models are designed for single tasks and rarely cover ripeness, diseases, pests and growth stages simultaneously; lightweight adaptation and field deployment verification for low-power edge chips represented by K230 remain insufficient; few systematic solutions combine edge detection results with multi-source data to achieve growth trend prediction and disease risk early warning. Therefore, this paper took greenhouse strawberries as the research object, adopted the YOLOv8 algorithm to establish four types of visual detection models, and trained the models based on continuously collected greenhouse growth images and public datasets. The lightweight model was transplanted to the K230 embedded chip, and automatic greenhouse inspection was realized with the support of the robot's autonomous inspection mechanism. Meanwhile, a VET-Net multimodal fusion architecture was built to integrate visual data, greenhouse temperature and humidity data, and hardware detection data, so as to realize 7-day crop growth situation and disease risk prediction.

2. Methodology

2.1. Detection Targets

According to the daily management requirements of strawberry planting, four core detection tasks were defined, covering the whole growth cycle from transplanting to harvesting:

(1) Strawberry ripeness detection: divided into three stages: green fruit stage, color transition stage, and mature stage, to guide harvesting schedule and avoid over-ripening loss.

(2) Strawberry disease identification: detection of powdery mildew and gray mold, the two most common and destructive diseases in greenhouse strawberry production.

(3) Strawberry pest identification: recognition of small pests including aphids, red spiders and thrips, which are difficult to detect manually at early stages.

(4) Growth stage classification: divided into flowering stage, young fruit stage, swelling stage and mature stage, to provide decision support for precision irrigation, fertilization and agricultural operations.

2.2. Image Preprocessing Methods

Problems such as uneven illumination, leaf occlusion and dust attachment in greenhouses can degrade image quality and increase the difficulty of model feature extraction. After comparative tests, the color feature segmentation method performed poorly in identifying diseases and pests with similar color to leaves and had a high false detection rate. Therefore, this paper adopted a preprocessing pipeline of "filter denoising - grayscale conversion - threshold segmentation":

(1) Bilateral filtering was applied to remove image noise, with kernel size set to 9 and standard deviations of both spatial and grayscale Gaussian functions set to 75. This algorithm smoothed the background noise while preserving edge textures of lesions, pests and fruits.

(2) The maximum value method was used to convert three-channel images into single-channel images to reduce computational load.

(3) Adaptive threshold segmentation was performed with BlockSize=27 and C=13, which enhanced the contrast between targets and leaf background and highlighted target contours. This preprocessing pipeline is consistent with the general image enhancement scheme for fine-grained plant disease detection, and significantly improves target recognizability under occlusion and low-light conditions.

To improve the scene robustness of the model, data augmentation was performed on the training set, including horizontal/vertical flipping, ±15° random rotation, 0.7-1.3 times random brightness adjustment, random cropping and scaling. A small amount of Gaussian noise was also added to simulate dust interference in greenhouses, expanding the training sample size to 1.5 times the original, simulating complex shooting scenarios and reducing the risk of overfitting.

2.3. YOLOv8 Algorithm Selection and Advantages

Relevant domestic reviews indicate that YOLOv8 has excellent small target detection performance and small model size, making it suitable for edge-side agricultural detection. This paper selected YOLOv8 to build the detection model for three main advantages:

(1) Optimized network structure. The C2f module replaced the original C3 module, which enriched gradient flow information by splitting feature branches. This design greatly improved the feature extraction ability for small targets such as tiny lesions and micro pests, and better adapted to the characteristics of small targets and frequent occlusion in agricultural scenarios.

(2) Balanced accuracy and speed. YOLOv8 optimized the anchor box allocation strategy and adopted the CIoU loss function to achieve a better balance between detection precision and inference efficiency. In this study, comparative tests were conducted on the self-built strawberry test set with a uniform input resolution of 640×640, using an NVIDIA RTX 3090 GPU as the test hardware. The experimental results showed that YOLOv8s reached an inference speed of 112 Frames Per Second (FPS) with a single-image inference latency of 8.9 ms, while YOLOv5s achieved 92 FPS with a latency of 10.9 ms under the same conditions. The inference speed of YOLOv8s was 21.7% higher than that of YOLOv5s, and its mAP@0.5 was 4.2 percentage points higher, confirming its performance advantage in strawberry multi-target detection scenarios.

(3) Convenient deployment. The model can run efficient local inference on the low-power embedded chip K230 without relying on cloud computing resources, which adapts well to scenarios with unstable network connectivity in rural greenhouses. Field test results showed

that at 640×640 input resolution, the lightweight YOLOv8s model achieved an inference speed of 18 FPS and a single-image latency of 55 ms on the K230 chip, fully satisfying the real-time demand of greenhouse autonomous inspection.

Wang H et al. verified the edge-side real-time detection performance of the lightweight YOLO model for facility tomato scenarios, and their experimental conclusions provide important reference for the embedded deployment scheme of this paper [9].

3. Model Construction and System Implementation

3.1. Image Annotation and Dataset Construction

Image annotation was meticulously performed prior to the model training phase, with LabelImg being employed to manually annotate all training images. To ensure a high level of annotation accuracy and consistency, unified annotation specifications were carefully formulated in advance: the annotation box should closely fit the target edge without including excessive background area; overlapping targets must be annotated separately without any omissions; furthermore, blurred or unrecognizable targets should not be annotated under any circumstances. These guidelines are essential to enhance the quality and reliability of the training data.

The total dataset contained 17,000 images, which were divided into training set, validation set and test set at a ratio of 8:1:1, with 13,600, 1,700 and 1,700 images respectively.

A detailed label system was established according to detection categories. There were three labels for ripeness: green fruit, color transition fruit and mature fruit; two labels for diseases: powdery mildew and gray mold as shown in Figure 1 and Figure 2; three labels for pests: aphids, red spiders and thrips as shown in Figure 3 and Figure 4; four labels for growth stages: flowering stage, young fruit stage, swelling stage and mature stage.

Figure 1. Annotated samples of strawberry diseases and ripeness stages.

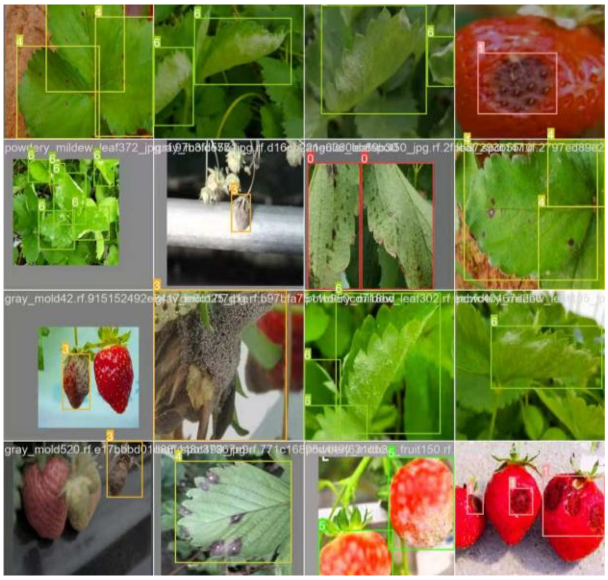

Figure 2. Strawberry diseases & ripeness.

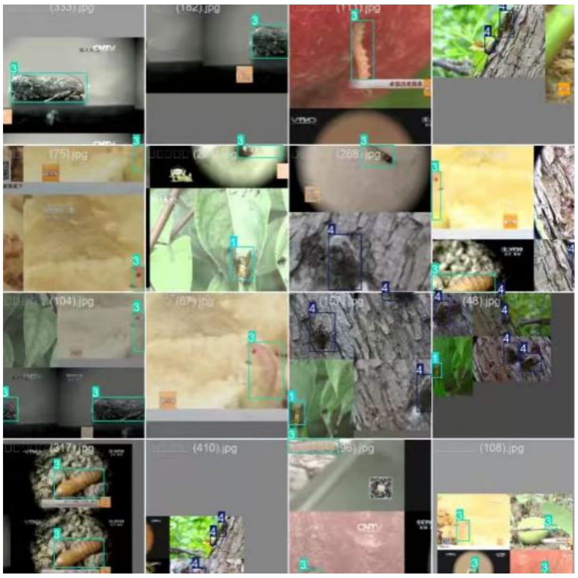

Figure 3. Strawberry pests.

Figure 4. Annotated samples of strawberry pests.

After annotation, YOLO-format label files and corresponding mapping files with original images were obtained, forming a standard dataset divided into training set, validation set and test set for subsequent model training as shown in Figure 5.

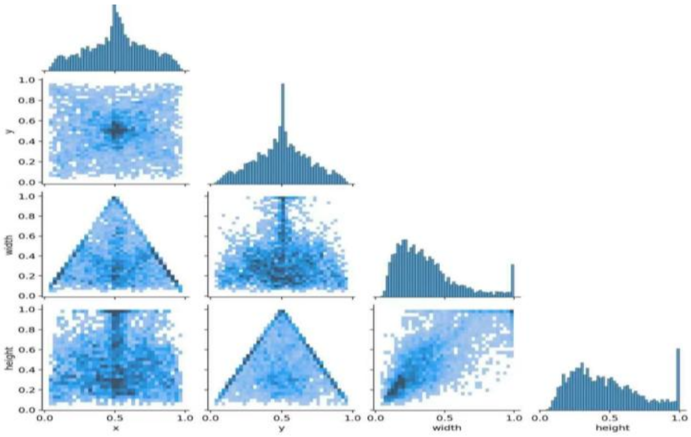

Figure 5. Dataset bounding box position & size distribution.

3.2. Model Training Environment and Parameter Settings

To accelerate model convergence and improve training performance, transfer learning was adopted to load pre-trained weights on public datasets for initialization. According to the dataset scale and hardware performance, core training parameters were determined: 300 epochs, batch size of 32, initial learning rate of 0.01, SGD (Stochastic Gradient Descent) optimizer, momentum factor of 0.937, and weight decay coefficient of 0.0005.

During training, the validation set was used to evaluate model accuracy after each epoch, and indicators such as precision, recall and mAP@0.5 were recorded. An early stopping strategy was adopted: training terminated when the validation set mAP did not improve for 10 consecutive epochs to prevent overfitting. Finally, the weight file with the best validation set accuracy was saved for embedded deployment.

3.3. Overall Implementation of the Edge-Cloud Collaborative System

The visual detection model was finally integrated into the greenhouse intelligent inspection system. The whole system adopted a four-layer architecture of "perception layer + edge layer + platform layer + application layer", realizing the complete closed loop from data collection to user application. Hou B et al. designed a complete software and hardware system for greenhouse inspection robots, and proposed the integrated development path of perception-decision-execution for facility agricultural robots [10].

The VET-Net (Vision-Environment-Terminal Network) multimodal fusion prediction module, as the core computing unit of the platform layer, adopted a three-branch feature-level fusion architecture with detailed structural parameters as follows:

(1) Visual feature branch: The input was the 128-dimensional global feature vector extracted from the backbone output of the edge-side YOLOv8 detection model. The branch consisted of 2 fully connected layers with dimensions of 256 and 128 respectively, each followed by a ReLU activation function, and finally output a 128-dimensional deep visual representation.

(2) Environmental temporal branch: The input was 5-dimensional continuous time-series data (air temperature, air humidity, light intensity, CO_2 concentration, soil moisture) collected by the perception layer for 7 consecutive days. The branch used 2 layers of LSTM network with a hidden layer dimension of 64 to extract temporal dependencies, and output a 64-dimensional environmental temporal feature.

(3) Hardware detection branch: The input was 32-dimensional basic crop state features (including target quantity statistics, disease lesion proportion, pest density, etc.) output by the K230 edge chip after local inference. The branch mapped the input to 64-dimensional features through 1 fully connected layer.

The features output by the three branches were concatenated for feature-level fusion, and then passed through two fusion fully connected layers (with dimensions of 256 and 128) with batch normalization processing. Finally, two parallel output heads were connected to respectively output the 7-day growth stage prediction results and the occurrence probability of common diseases.

The system data transmission flow was as follows: the robot took photos at regular intervals, and uploaded images, detection results and environmental sensing data to the cloud via a 4G cellular network after local identification; after the cloud completed data storage and prediction calculation via VET-Net, the results were pushed to the Mini Program. Field tests showed that the end-to-end delay from data collection completion to full display on the Mini Program was controlled within 2 seconds, which could meet the demand for real-time early warning of pests and diseases. Huang W et al. pointed out in their research on strawberry pest and disease IoT monitoring system that stable low-latency data transmission is the core premise for the implementation of greenhouse remote monitoring and early warning functions [11].

4. Experimental Results and Data Analysis

4.1. Accuracy Analysis of Four Types of Visual Target Detection

After model training, an independent test set was used for performance testing, and the identification accuracy of the four types of targets was counted. The detailed test results are shown in Table 1:

Table 1. Statistics of identification test results for four types of targets.

Test category	**Test set samples (Images)**	**Correctly identified samples (Images)**	**Identification accuracy**
Strawberry ripeness	400	379	94.8%
Disease identification	400	380	94.9%
Pest identification	400	379	94.7%
Growth stage	300	282	94.1%

According to the test results, the identification accuracy of the four target detectors integrated in a single model exceeded 94%, balancing the richness of detection functions and identification accuracy. Disease identification achieved the highest accuracy of 94.9%, mainly because the lesions of powdery mildew and gray mold were distinct and exhibited significant color difference from the leaf background, making it easy for the model to extract features as shown in Figure 6. Xia S et al. carried out strawberry leaf disease detection based on improved YOLOv8n, with the baseline model achieving an average identification accuracy of 92.3%. The model in this paper further improved the detection accuracy on this basis [12]. The improved YOLOv8 in this paper achieved higher accuracy, which aligns with the technology development trend.

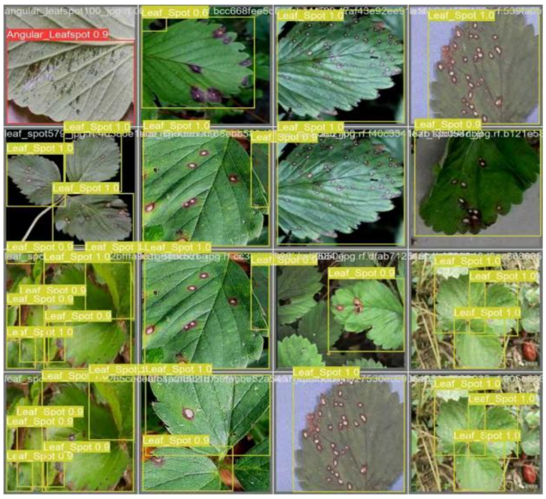

Figure 6. Strawberry leaf disease model.

To further analyze the classification performance of various disease categories in detail, the confusion matrix for these disease categories on the test set was plotted, as illustrated in Figure 7. Both powdery mildew and gray mold exhibited a sufficient number of identification samples, along with a notably low inter-category false detection rate. Only a small number of samples pertaining to fruit powdery mildew and gray mold were slightly confused, which can be attributed to the high visual similarity observed in the lesions on the surfaces of the fruit. Overall, the classification accuracy achieved met the stringent requirements necessary for effective field detection and practical applications in agriculture.

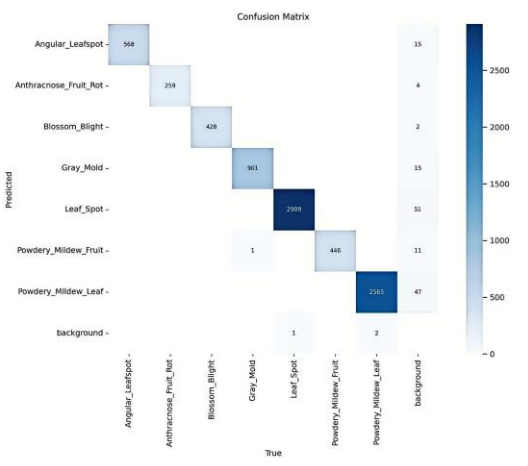

Figure 7. Strawberry disease detection confusion matrix.

The accuracy of strawberry ripeness detection reached 94.8%, and the precision curve of each category varying with confidence threshold is shown in Figure 8. The precision of all categories can reach 1.00 at the confidence threshold of 0.955, indicating that the model output has high overall confidence and very few false detections under high confidence, which ensures the reliability of on-site greenhouse detection results. Bao Y et al. carried out multi-stage strawberry ripeness classification detection based on deep learning, and the baseline identification accuracy of their model is close to the results of this paper [13]; Zhou K et al.'s research on YOLO-based detection of small farmland pests shows that the average identification accuracy of mainstream algorithms in such scenarios is about 93%, consistent with the pest detection results of this paper [14], while the model in this paper supports multi-category detection with better comprehensive performance.

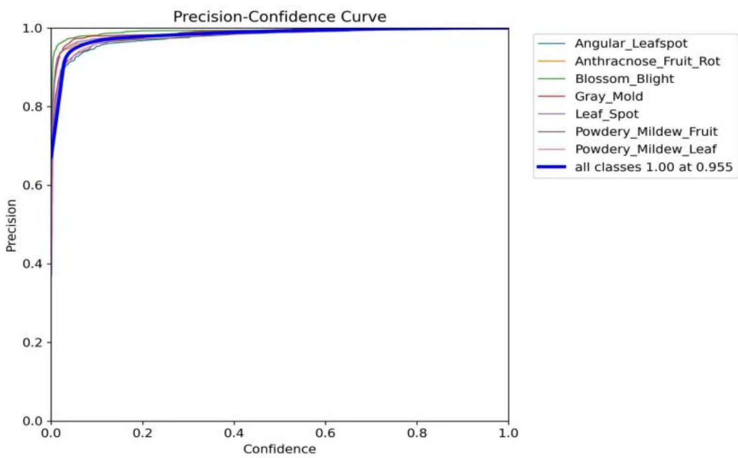

Figure 8. Per-class precision-confidence curves of strawberry detection targets.

4.2. Effect of Multimodal Fusion Growth Trend Prediction

Visual data alone can only reflect the current state of crops, but cannot predict growth trends. This paper constructed a VET-Net multimodal fusion prediction structure: visual features of images were extracted through CNN, temporal features of environmental data such as temperature, humidity and illumination were extracted through LSTM, and basic crop features output from the K230 end were fused. Finally, 7-day crop growth prediction and disease risk assessment results were output. Du Y et al. confirmed in their research on multimodal diagnosis of crop diseases that the multimodal scheme integrating visual features and environmental data can effectively compensate for the limitation that single-source image data cannot reflect the dynamic law of crop growth [15].

To verify the improvement effect of multimodal fusion, four controlled experiments were set up to test the growth trend prediction accuracy under different input information, as shown in Table 2:

Table 2. Growth trend prediction accuracy under different input combinations.

Model input combination	**Prediction accuracy**
Image visual data only	80.00%
Vision + greenhouse environmental data	89.00%
Vision + K230 hardware detection data	92.30%
Vision + environment + hardware detection (VET-Net)	98.62%

The experimental results showed that prediction relying only on visual data performed the worst, as it cannot reflect the impact of environmental changes on crop growth. With the addition of environmental data and hardware detection data, the prediction accuracy gradually improved. The VET-Net prediction model built by fusing all three types of data achieved an accuracy of 98.62%, which was 18.62 percentage points higher than the scheme using only visual data. Fu J et al. sorted out the fusion technology of greenhouse environment regulation and visual monitoring, and proposed that multimodal collaboration is the core scheme for crop growth trend prediction, which is consistent with the experimental conclusion of this paper [16].

The model has the capability to predict the occurrence probability of diseases, such as powdery mildew, with a remarkable one-week advance notice. Additionally, it provides corresponding improvement suggestions that include effective water and fertilizer regulation, enhanced ventilation, and preventive disinfection measures. Based on the comprehensive data from base pilot studies, the implementation of this advanced prediction scheme has led to a significant improvement in the timeliness rate of disease prevention and control efforts. As a result, strawberry planting costs have been notably reduced by 18%, while the yield per unit area has impressively increased by 10%. This innovative approach not only optimizes resource management but also enhances overall crop health and productivity.

4.3. Performance verification of inspection robot supporting system

Aiming at complex obstacles such as supports, water pipes and fallen leaves in greenhouses, a vision + ultrasonic fusion obstacle avoidance scheme was adopted and compared with single-sensor schemes. The results showed that the fusion scheme achieved an obstacle avoidance success rate of 94% with an average response time of 305ms; single vision-based obstacle avoidance had a success rate of 78% and an average response time of 420ms, with poor performance in identifying transparent obstacles and fine branches; single ultrasonic obstacle avoidance had a success rate of 82% and an average response time of 350ms, which cannot distinguish obstacle types and is prone to invalid detours. The fusion scheme is more adaptable to the complex greenhouse environment and can ensure the safety of robot autonomous inspection. Ji L et al.'s experiments on the robot obstacle avoidance system for greenhouse strawberry scenarios showed that the vision and multi-sensor fusion scheme has significantly better obstacle avoidance performance than single-sensor schemes, which is consistent with the experimental conclusion of this paper [17].

5. Conclusion

Addressing the significant industry pain points of low manual inspection efficiency and delayed detection of pests and diseases in greenhouse strawberry cultivation, this paper establishes a comprehensive multi-target visual detection model based on YOLOv8. This advanced model effectively facilitates the simultaneous detection of four critical types of targets: strawberry ripeness, various diseases, prevalent pests, and different growth stages. Furthermore, multimodal fusion technology has been employed to accurately predict crop growth trends, which was thoroughly verified through rigorous laboratory tests and practical applications in the field. The conclusions drawn from this study are as follows:

The four-type target detection model based on YOLOv8 achieved over 94% identification accuracy for all four types of targets after 300 epochs of training on a dataset of 17,000 images. The lightweight model can be stably deployed on the embedded end to meet the requirements of real-time greenhouse inspection. The VET-Net multimodal fusion architecture integrated visual, environmental and hardware data, and achieved 98.62% accuracy for 7-day crop growth trend prediction, enabling early detection of disease risks and supporting farmers in refined management. After the whole system is implemented with solar-powered autonomous inspection robots, the inspection efficiency was 6 times that of manual inspection, and planting cost was reduced by 18%, showing good cost reduction and efficiency improvement effects as well as promotion value.

References

[1] Ge, L., Cheng, Y., Qiu, L., Wang, X., Zhai, C., & Ding, X. (2024). Application progress of agricultural internet of things technology in greenhouse production. *Journal of Zhejiang Agricultural Sciences*, 65(1), 242-248.

[2] Xue, X., Liu, P., & Zhou, W. (2024). Real-time fine-grained plant disease detection based on improved YOLOv8 algorithm. *Journal of Chinese Agricultural Mechanization*, 45(5), 188-195.

[3] Peng, W., & Luo, J. (2025). Identification method of tea leaf pests and diseases in hilly areas based on improved YOLOv7. *Journal of Agricultural Mechanization Research*, 47(12), 45-52.

[4] Chen, D., & Kuang, C. (2024). Detection method of tea pests and diseases based on YOLOv5s. *Fujian Tea*, 46(3), 32-34.

[5] Xu, Y., & Chen, L. (2024). Tomato leaf disease detection algorithm based on improved YOLOv8. *Jiangsu Agricultural Sciences*, 52(17), 192-200.

[6] Mao, S., & Wang, W. (2024). Research review of YOLO series object detection algorithms based on deep learning. *Journal of Yan'an University (Natural Science Edition)*, 43(2), 88-95.

[7] Wan, Y., & Yan, Z. (2024). Research review of crop pest and disease identification based on YOLO algorithm. *Tropical Agricultural Engineering*, 48(1), 25-28.

[8] Zhou, Y., Liu, D., & Zhou, Y. (2024). Detection of crop disease leaves based on multimodal feature alignment. *Journal of Chinese Agricultural Mechanization*, 45(7), 132-138.

[9] Wang, H., Sun, L., Li, X., Liu, H., Wang, G., & Lan, Y. (2024). Detection method of tomato leaf pests and diseases based on improved YOLOv7-tiny. *Transactions of the Chinese Society of Agricultural Engineering*, 40(10), 194-202.

[10] Hou, B., Li, X., Mou, X., & Yao, H. (2024). Design and experiment of greenhouse inspection robot for cherry tomatoes. *Journal of Chinese Agricultural Mechanization*, 45(1), 86-92.

[11] Huang, W., Zhou, X., Wang, C., Feng, Z., & Chen, Z. (2024). Design of IoT strawberry pest and disease monitoring system based on YOLOv5. *Journal of Lanzhou Institute of Technology*, 31(6), 72-76.

[12] Xia, S., Ni, M., Luo, Y., He, Y., & Zhao, T. (2025). Strawberry leaf disease detection method based on improved YOLOv8n. *Jiangsu Journal of Agricultural Sciences*, 41(4), 664-675.

[13] Bao, Y., Zhang, Y., Fan, C., & Liu, S. (2025). Strawberry ripeness detection method based on deep learning. *Jiangsu Agricultural Sciences*, 53(5), 89-94.

[14] Zhou, K., Liu, X., & Zheng, T. (2024). Research on small target multi-category farmland pest detection algorithm based on improved YOLOv5. *Journal of Chinese Agricultural*

Mechanization, 45(6), 235-241.

[15] Du, Y., Huang, Y., Du, P., Gao, X., Wu, M., & Yang, Y. (2023). Research on intelligent collaborative diagnosis model of tomato gray mold based on multimodal fusion technology. *Journal of Chinese Agricultural Mechanization*, 44(11), 115-122.

[16] Fu, J., He, X., Wang, X., Liu, J., Zhu, G., & Li, M. (2024). Status analysis and development research of greenhouse environment control methods. *Journal of Chinese Agricultural Mechanization*, 45(4), 97-103.

[17] Ji, L., Liu, Y., Gao, J., & Wu, D. (2023). Design and experiment of greenhouse strawberry picking robot. *Journal of Chinese Agricultural Mechanization*, 44(1), 192-198.