A Self-Prompted YOLOv11–SAM 2 Pipeline for Automatic Plant Disease Detection and Segmentation

Balkis Tej
Automatic Signal and Image Processing Research Laboratory (LR13ES13), National Engineering School of Monastir, University of Monastir, Monastir, Tunisia
balkis.tej@enim.u-monastir.tn (corresponding author)

Soulef Bouaafia
Laboratory of Condensed Matter and Nanoscience (LR11ES40), Faculty of Sciences of Monastir, University of Monastir, Tunisia | Higher Institute of Applied Sciences and Technology of Kairouan, University of Kairouan, Kairouan, Tunisia
soulef.bouaafia@fsm.rnu.tn

Mohamed Ali Hajjaji
Research Laboratory in Algebra Numbers Theory and Intelligent Systems (RLANTIS), University of Monastir, Tunisia | Higher Institute of Applied Sciences and Technology of Sousse, University of Sousse, Sousse, Tunisia
mohamedali.hajjaji@issatso.rnu.tn

Abdellatif Mtibaa
Systems Integration & Emerging Energies Laboratory (LR21ES14), National Engineering School of Sfax, University of Sfax, Sfax, Tunisia
abdellatif.mtibaa@enim.rnu.tn

Received: 11 November 2025 | Revised: 25 November 2025 and 4 December 2025 | Accepted: 7 December 2025

 DOI: https://doi.org/10.48084/etasr.16198

ABSTRACT

Plant disease detection and segmentation are essential to maintaining healthy crops and improving agricultural productivity. Using Artificial Intelligence (AI) for this task enables farmers and researchers to identify diseases early and take preventive action. However, accurately segmenting diseased regions remains challenging, as most methods require a large number of labeled images or manual guidance to train the model. This limits their scalability and practical use in real-world agricultural settings. To address this issue, we propose the YOLO_SAM model for plant disease detection and segmentation, using self-prompted mask generation. Experiments were conducted on a self-generated dataset containing samples of different leaf diseases. The YOLO11 model was first used to detect infected regions, and its bounding boxes were automatically passed to the Segment Anything Model (SAM), which generated detailed segmentation masks without manual input. Experimental results showed that YOLOv11s achieved the best performance with a mean Average Precision (mAP) of 0.845, outperforming YOLOv11n (0.825), YOLOv11l (0.807), and YOLOv11m (0.764). Based on this superior performance, YOLOv11s was chosen as the prompt generator for SAM2, enabling more accurate and reliable segmentation of disease regions. This combined approach enabled the model to generate precise lesion masks without manual prompting, allowing clear boundary extraction even for small or irregular disease spots.

Keywords-plant disease detection; segmentation; YOLOv11; segment anything model

I. INTRODUCTION

One of the most pressing problems that farmers face today is plant disease. These diseases can severely reduce agricultural yields, posing a major threat to both food security and the growth of sustainable farming [1]. The Food and Agriculture Organization (FAO) estimates that these diseases cause the world to lose between 20% and 40% of its food each year [2]. Therefore, the ability to detect and locate disease quickly and accurately is essential to maintaining productive and resilient agricultural systems. Traditional plant disease control relies on manual observation by farmers or agricultural experts. This process takes a lot of time, depends on personal opinions, and is limited by human expertise and environmental factors. Visual inspection can also be insufficient when assessing large agricultural regions [3]. To address these problems, computer vision and AI have become powerful tools for automating the identification of plant diseases. By analyzing leaf images, AI algorithms can detect, classify, and localize diseased regions with high precision and speed [4]. Deep learning models, especially Convolutional Neural Networks (CNNs), have transformed agricultural image analysis over the past few years. Among these models, the YOLO (You Only Look Once) family has become one of the most widely used approaches for detecting plant diseases [5, 6]. At the same time, image segmentation also plays an important role, helping to clearly identify infected areas on leaves and estimate disease severity [7]. Traditional segmentation networks such as U-Net [8] and DeepLab [9] have achieved good results; however, their adaptability and generality often decrease when applied to diverse plant disease scenarios [10]. More recently, the SAM, trained on a wide variety of images, has demonstrated strong generalization [11], making it a promising tool for plant disease segmentation. Several studies have examined the use of foundation models for agricultural image segmentation.

Authors in [12] proposed the MoE-SAM model, an improved version of the Segment Anything Model, for segmenting plant diseases. Their approach integrates sparse attention and mixture-of-experts techniques, improving segmentation accuracy while reducing computational costs. Authors in [13] combine the SAM with a Fully Convolutional Data Description (FCDD) network to improve segmentation accuracy in real-world conditions. The model isolates healthy and diseased leaves from complex backgrounds, achieving over 10% higher accuracy than conventional CNN-based methods. In [14], the authors developed the SAM-IE model, an enhanced segmentation model that integrates image preprocessing techniques, such as CLAHE and bilateral filtering, with SAM prompts to segment cucumber leaf infections more accurately. The authors in [15] employed the SAM model in conjunction with YOLOv8 to detect and segment tomato leaves from plants within a real-time environment.

While these studies have shown significant progress in applying YOLO models for plant disease detection, the models still struggle to generalize in real agricultural environments due to variations in lighting, background complexity, and the presence of small or subtle lesions. This limitation results in bounding boxes that do not always capture fine disease boundaries [16].

Likewise, although SAM-based methods have shown promising results for plant disease detection and segmentation, they often require large amounts of annotated data and perform poorly on small, overlapping, or low-contrast lesions [17]. To overcome this limitation, this study presents a pipeline architecture that combines YOLOv11 for plant leaf disease detection and the SAM 2 model for segmentation. In the proposed approach, the bounding boxes predicted by YOLOv11 serve as spatial prompts for SAM 2, helping the model focus on diseased areas and generate refined segmentation masks. The main contribution of this work is:

- Application of multiple YOLOv11 variants for plant leaf disease detection, enabling a comparative evaluation to identify the most suitable model for precise localization of diseased regions.
- Integration of YOLOv11 and SAM2 by using the bounding boxes generated by the best-performing YOLOv11 variant as automatic prompts for a fine-tuned SAM2 model to segment plant leaf diseases.
- Development of a self-generated dataset comprising images of leaf diseases impacting tomato and pepper crops in Tunisia.

II. MATERIALS AND METHODS

A. Dataset

1) Dataset Description

For this study, a self-generated dataset was assembled [18] comprising healthy and diseased tomato and pepper leaf specimens. It includes 522 images organized into six classes: a) healthy leaves, b) Tomato Yellow Leaf Curl Virus (TYLCV), c) nutrient deficiencies, d) oidium, e) leaf miner, and f) mildew, as shown in Figure 1. The images were captured with a camera phone in various tomato and pepper greenhouses in Monastir, Tunisia, under the supervision of an agricultural engineer.

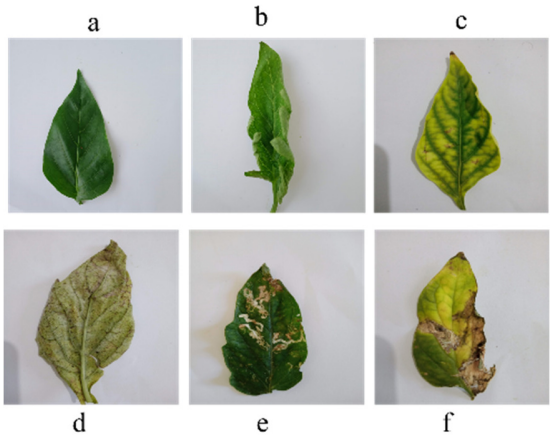

Fig. 1. Example of images: (a) healthy, (b) TYLCV, (c) nutrient deficiencies, (d) oiduim, (e) leaf miner, (f) mildew.

2) Dataset Pre-Processing

The data pre-processing phase comprises two primary stages:

- Data annotation: The dataset images were manually annotated using the Roboflow tool to outline infected areas on leaves. Each diseased region was marked with a bounding box and assigned to its respective disease category. The annotations were subsequently exported as YOLO-formatted text files, containing the class ID and normalized coordinates (x, y, width, height) for each object in the image.
- Data augmentation: Owing to the limited size of our dataset and the irregular class distribution, different data augmentation techniques were applied to the training data to improve model generalization. The transformations included image rotations, horizontal and vertical flips, Gaussian blur, and the addition of random noise. Adjustments to color properties, such as saturation, exposure, and brightness, were also applied to simulate lighting and image-quality variations. The total number of images after applying data augmentation techniques increased to 1352 images.

B. Yolov11 Model

The YOLOv11 model was used in this study to detect and generate bounding boxes around diseased areas on plant leaves. This version retains the strengths of earlier models and introduces several architectural improvements to enhance detection accuracy and computational efficiency [19]. YOLOv11 comes in several configurations, including nano (n), small (s), medium (m), and large (l), each balancing accuracy, speed, and computational cost differently [20]. The architecture of YOLOv11 introduces several improvements to its backbone and neck to enhance feature extraction, resulting in better detection accuracy in complex scenarios (Figure 2). The main changes include replacing the CF2 module with C3K2, adding a C2PSA block after the SPPF layer, and adopting the YOLOv10 head structure. In addition, it uses depth-wise separable convolutions, which help reduce unnecessary computations, making the model faster and more efficient overall [21].

C. Segment Anything Model (SAM v2)

SAM 2 [22], which builds on Meta's original Segment Anything Model, delivers significant improvements in object area segmentation for both images and videos. It maintains SAM's strength in generating high-quality masks from simple prompts and goes further by enabling real-time segmentation with minimal human interaction, making it highly useful for applications such as plant disease analysis. The architecture of the SAM 2 model comprises four main blocks [23], as shown in Figure 3. The image encoder uses a transformer architecture to learn important image features, enabling the model to understand visual content. The prompt encoder accepts user inputs, such as points, boxes, or masks, to guide segmentation and focus on specific objects. The memory component includes a memory encoder, a memory bank, and an attention module that retains information from previous frames, enabling the model to track objects over time. Finally, the mask decoder generates final segmentation masks using both image features and prompts, while incorporating memory information to improve accuracy across frames.

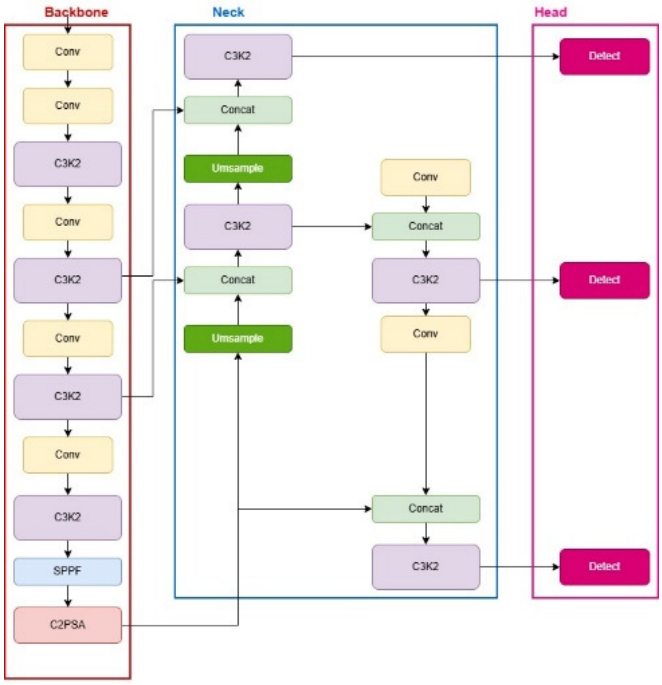

Fig. 2. Yolo11 architecture.

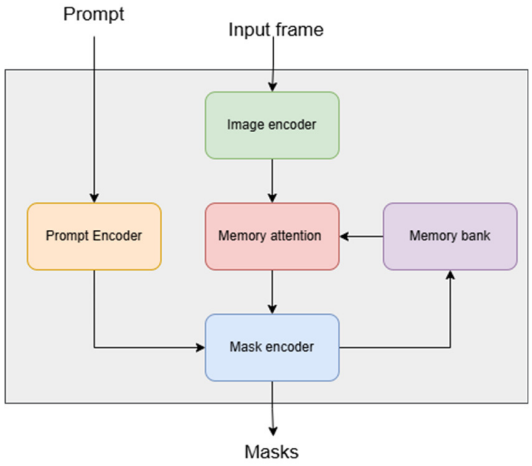

Fig. 3. SAM architecture.

D. The Proposed Pipeline Architecture

The architecture of our proposed model for plant disease detection and segmentation is shown in Figure 4. It integrates YOLOv11's detection capability with SAM2's segmentation functionality in a unified pipeline. The process begins with an input leaf image, which is passed through the YOLOv11 model to detect and localize disease areas by generating bounding boxes around them. In this work, the SAM2 model was used in a zero-shot configuration, with the pretrained sam2_b.pt

weights applied directly. The YOLOv11 model's bounding boxes were used as prompts to guide SAM2 in generating segmentation masks. Inside SAM 2, the image passes through an image encoder that extracts visual features, followed by a memory attention module and a mask decoder, which produce detailed segmentation masks of the infected regions. The memory encoder and memory bank help retain contextual information, improving consistency across predictions. This self-prompted setup enables the system to automatically detect and segment diseased leaf regions without manual annotation, producing both colored and binary masks that precisely highlight the affected zones.

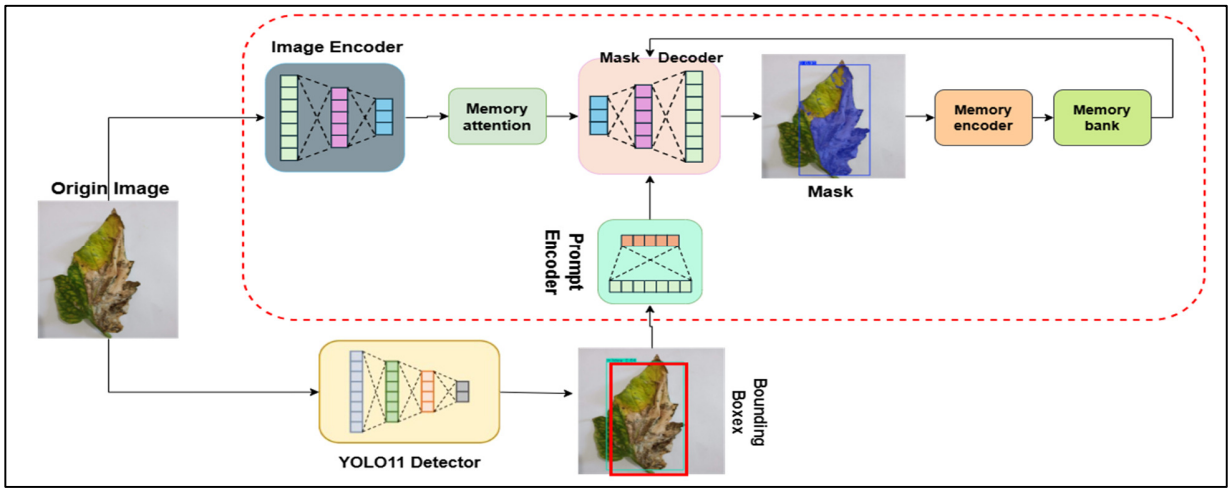

Fig. 4. The proposed architecture YOLO-SAM.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

We conducted the training phase in the Google Colab Pro environment, which provided a practical, efficient cloud-based platform for deep learning experimentation. Colab Pro enabled us to access high-performance Graphics Processing Unit (GPU) resources without dedicated local hardware, ensuring stable computation and accelerated model training. The dataset was split into two subsets: 80% for training and 20% for validation. The hyperparameters and training configurations used to optimize all YOLOv11 variants are summarized in Table I.

TABLE I. TRAINING HYPERPARAMETERS

Parameter	Value
Input image size	640 x 640
Batch size	16
Number of epochs	100
Optimizer	SGD
Momentum	0.937
Learning rate	0.01
Weight decay	0.0005

B. Detection Results of YOLOv11 Model

To decide which YOLOv11 variant to use as the prompt generator for the SAM 2 model, we compared their detection performance. Table II shows this comparison based on the main evaluation metrics [24]:

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (1)$$

$$\text{AP} = \int_0^1 P.Rdr \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F1} - \text{score} = 2 \times \frac{(Precision \times Recall)}{Precision+Recall} \quad (5)$$

TABLE II. TABLE TYPE STYLES

Model	mAP	Precision (P)	Recall (R)	F1-score
YOLOv11n	0.825	0.868	0.733	0.78
YOLOv11s	0.845	0.857	0.804	0.829
YOLOv11m	0.764	0.722	0.763	0.741
YOLOv11l	0.807	0.779	0.79	0.784

As illustrated, YOLOv11s achieved the highest mAP of 0.845, indicating that it performs best overall in detecting diseased areas. It also has a good balance between precision (0.857) and recall (0.804), indicating it can detect most infected regions while keeping false detections low. The YOLOv11n version came close but had slightly lower recall, missing some smaller disease spots. The larger models, YOLOv11m and YOLOv11l, did not perform as well, likely because of their higher complexity relative to the dataset size. Overall, YOLOv11s performed best, so it was chosen to provide the detection prompts for the SAM 2 model. Figure 5 illustrates how the YOLOv11s model performed during training and validation over 100 epochs. We can see that the box, classification, and DFL losses gradually decrease, indicating that the model continues to improve at correctly locating and classifying diseased areas. At the same time, precision, recall, and mAP continue to increase across epochs, indicating that the model is becoming more accurate and consistent. Toward the end of the training, the curves stabilize, indicating the model has reached good convergence without overfitting.

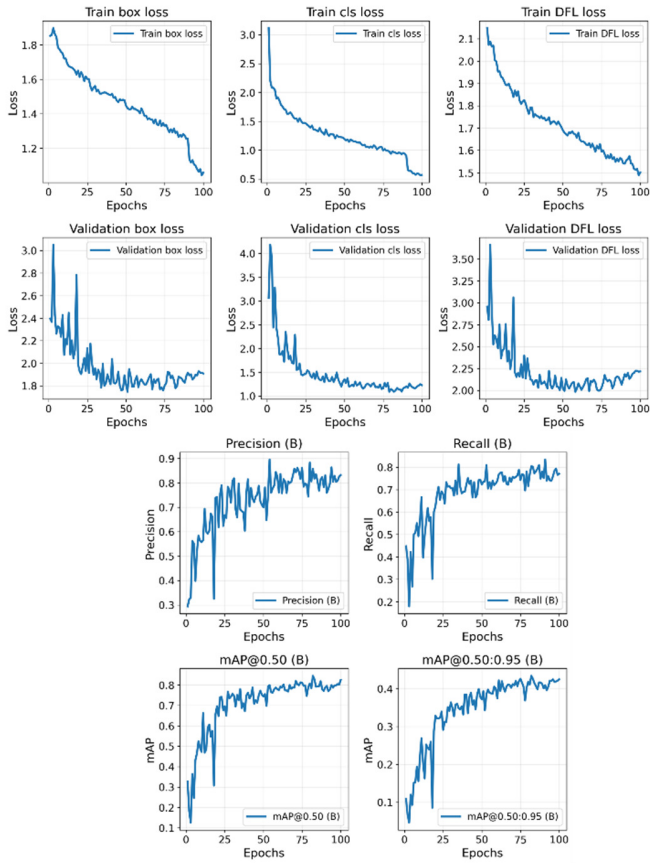

Fig. 5. YOLOv11s training results.

Figure 6 shows the PR curves for the YOLOv11s model. It shows how well YOLOv11s detects each type of leaf disease. The curves were generated using an IoU threshold of 0.5 (i.e., mAP@0.5), which is the evaluation criterion used in this study. Overall, the model achieved a mean average precision of 0.845, indicating strong detection across all classes. The nutrient deficiency (0.995) and TYLCV (0.961) classes achieved the highest precision and recall, indicating they are the easiest to identify with minimal false detections. Healthy leaves (0.946) also performed well, indicating the model can reliably distinguish them from diseased leaves. By contrast, leaf miner (0.694), mildew (0.736), and oidium (0.738) show lower precision and recall, indicating these diseases are more challenging to detect.

C. The SAM and YOLO Segmentation Results

Figure 7 shows the performance of the proposed pipeline architecture on various image examples from the dataset. The first column shows the original input images, which include different types of leaf diseases: the first image shows a leaf with mildew, the second with oidium, the third with leaf miner, and the last with Tomato Yellow Leaf Curl Virus (TYLCV). The second column shows the images after disease detection with the YOLOv11s model, with bounding boxes around the diseased areas. The third column shows the output of the segmentation process, generated with SAM2 using YOLO's bounding boxes as prompts.

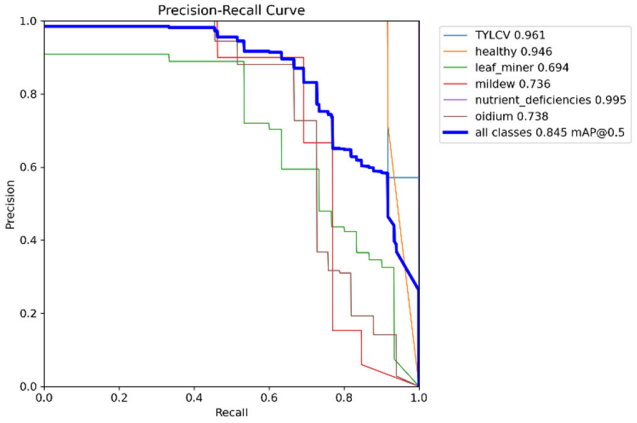

Fig. 6. YOLOv11s PR-curves.

Fig. 7. YOLO+SAM segmentation results.

We observe that YOLOv11 and SAM2 work effectively together to detect and isolate diseased regions. The model performs well even when the leaf textures are complex or overlapping, as seen in the first and third images. For smaller or less visible infections, such as in the second image, SAM2 can still generate masks that accurately cover the affected areas. These results demonstrate that the YOLOv11–SAM2 pipeline can go beyond basic detection to achieve a detailed,

pixel-level understanding of leaf disease regions, which is useful for monitoring and early diagnosis in agriculture.

IV. CONCLUSION

This paper presents a self-prompted pipeline architecture that combines YOLOv11 for object detection and SAM2 for image segmentation to effectively identify and segment plant leaf diseases. The proposed approach automatically generates segmentation masks without manual intervention. A comparison of YOLOv11 variants showed that YOLOv11s achieved the best overall performance, making them the most suitable choice for providing bounding-box prompts to SAM2. Experimental results on a self-generated dataset of tomato and pepper leaves demonstrated that integrating YOLOv11 and SAM2 provides accurate localization and fine segmentation of diseased regions, even under complex visual conditions. Future research may expand the dataset by incorporating real-world field images to improve robustness. Additionally, it should evaluate the pipeline across a broader range of crop species and disease categories. Investigating optimized model variants for real-time inference on edge devices is also recommended. Furthermore, exploring multimodal sensing techniques, such as hyperspectral and thermal data, could significantly improve early disease detection capabilities.

REFERENCES

[1] Z. Ge, X. Fan, J. Zhang, and S. Jin, "SegPPD-FS: Segmenting plant pests and diseases in the wild using few-shot learning," *Plant Phenomics*, vol. 7, no. 4, Dec. 2025, Art. no. 100121, https://doi.org/10.1016/j.plaphe.2025.100121.

[2] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson, "The global burden of pathogens and pests on major food crops," *Nature Ecology & Evolution*, vol. 3, no. 3, pp. 430–439, Mar. 2019, https://doi.org/10.1038/s41559-018-0793-y.

[3] B. Tej, S. Bouaafia, M. A. Hajjaji, and A. Mtibaa, "AI-based smart agriculture 4.0 system for plant diseases detection in Tunisia," *Signal, Image and Video Processing*, vol. 18, no. 1, pp. 97–111, Aug. 2024, https://doi.org/10.1007/s11760-024-03134-z.

[4] S. Zhang and C. Zhang, "Modified U-Net for plant diseased leaf image segmentation," *Computers and Electronics in Agriculture*, vol. 204, Jan. 2023, Art. no. 107511, https://doi.org/10.1016/j.compag.2022.107511.

[5] Y. Alhwaiti, M. Khan, M. Asim, M. H. Siddiqi, M. Ishaq, and M. Alruwaili, "Leveraging YOLO deep learning models to enhance plant disease identification," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, Art. no. 7969, https://doi.org/10.1038/s41598-025-92143-0.

[6] Y. Meng, J. Zhan, K. Li, F. Yan, and L. Zhang, "A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, Art. no. 6016, https://doi.org/10.1038/s41598-025-88399-1.

[7] M. Shoaib *et al.*, "Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease," *Frontiers in Plant Science*, vol. 13, Oct. 2022, Art. no. 1031748, https://doi.org/10.3389/fpls.2022.1031748.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

[9] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation." arXiv, Aug. 22, 2018, https://doi.org/10.48550/arXiv.1802.02611.

[10] J. Li, Q. Feng, J. Zhang, and S. Yang, "EMSAM: enhanced multi-scale segment anything model for leaf disease segmentation," *Frontiers in Plant Science*, vol. 16, Mar. 2025, Art. no. 1564079, https://doi.org/10.3389/fpls.2025.1564079.

[11] F. Alfiaturrohmah and S. Sudianto, "A Segment Anything Model for Melon Pruning Based on Diameter," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 26632–26639, Oct. 2025, https://doi.org/10.48084/etasr.12207.

[12] B. Zhao *et al.*, "Sparse-MoE-SAM: A Lightweight Framework Integrating MoE and SAM with a Sparse Attention Mechanism for Plant Disease Segmentation in Resource-Constrained Environments," *Plants*, vol. 14, no. 17, Aug. 2025, Art. no. 2634, https://doi.org/10.3390/plants14172634.

[13] E. Moupojou, F. Retraint, H. Tapamo, M. Nkenlifack, C. Kacfah, and A. Tagne, "Segment Anything Model and Fully Convolutional Data Description for Plant Multi-Disease Detection on Field Images," *IEEE Access*, vol. 12, pp. 102592–102605, 2024, https://doi.org/10.1109/ACCESS.2024.3433495.

[14] R.-W. Bello, P. A. Owolawi, E. A. V. Wyk, and C. Tu, "SAM-IE: SAM-enabled image enhancement for segmentation of infected cucumber leaves," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 2, pp. 824–832, Mar. 2025, https://doi.org/10.53894/ijirss.v8i2.5328.

[15] S. U. Islam, G. Ferraioli, and V. Pascazio, "Tomato Leaf Detection, Segmentation, and Extraction in Real-Time Environment for Accurate Disease Detection," *AgriEngineering*, vol. 7, no. 4, Apr. 2025, Art. no. 120, https://doi.org/10.3390/agriengineering7040120.

[16] R. Kaur *et al.*, "YOLO-LeafNet: a robust deep learning framework for multispecies plant disease detection with data augmentation," *Scientific Reports*, vol. 15, no. 1, Aug. 2025, Art. no. 28513, https://doi.org/10.1038/s41598-025-14021-z.

[17] J. Reddy *et al.*, "Cotton Yield Prediction via UAV-Based Cotton Boll Image Segmentation Using YOLO Model and Segment Anything Model (SAM)," *Remote Sensing*, vol. 16, no. 23, Jan. 2024, Art. no. 4346, https://doi.org/10.3390/rs16234346.

[18] "Leaf_disease_dataset Object Detection Model by balkis," *Roboflow*. https://universe.roboflow.com/balkis/leaf_disease_dataset-ncmi5.

[19] X. Zhang, R. Liu, R. Liu, D. Yang, and J. Yang, "A non-contact rabbit temperature detection model based on improved YOLO11," *Smart Agricultural Technology*, vol. 12, Dec. 2025, Art. no. 101350, https://doi.org/10.1016/j.atech.2025.101350.

[20] S. Teboulbi, S. Messaoud, M. A. Hajjaji, M. Atri, and A. Mtibaa, "Fine Tuned YOLOv11-Based Road Sign Detection," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24950–24956, Aug. 2025, https://doi.org/10.48084/etasr.11529.

[21] J. Meng *et al.*, "P-YOLO11: An improved lightweight model for accurate detection of declining trees in poplar plantations," *Smart Agricultural Technology*, vol. 12, Dec. 2025, Art. no. 101454, https://doi.org/10.1016/j.atech.2025.101454.

[22] N. Ravi *et al.*, "SAM 2: Segment Anything in Images and Videos." arXiv, Oct. 28, 2024, https://doi.org/10.48550/arXiv.2408.00714.

[23] Ultralytics, "SAM 2: Segment Anything Model 2." https://docs.ultralytics.com/models/sam-2/.

[24] J. He, Y. Ren, W. Li, and W. Fu, "YOLOv11-RCDWD: A New Efficient Model for Detecting Maize Leaf Diseases Based on the Improved YOLOv11," *Applied Sciences*, vol. 15, no. 8, Jan. 2025, Art. no. 4535, https://doi.org/10.3390/app15084535.