Visual Prompt and Reference-Guided Plant Health Detection and Classification

using a two-stage pipeline with SAM2 and a fine-tuned classifier

Athul Dev, Master Computer Vision and Data Science

Supervisors: Willem Dijkstra

Winter 2025

Introduction

- Manual plant health assessments are error-prone and labor-intensive. [1]
- PVY detection via hyperspectral imaging is costly. [2]
- This research focuses on:
  - Decoupled pipeline combining prompt-based segmentation followed by independently optimized classification.
  - RGB-Based classification leveraging ViT-B/16, ResNet-18, and EfficientNet-B0 with reference image guidance via feature embeddings for PVY classification.

Abstract

This study addresses Potato Virus Y (PVY) detection by combining SAM2-based segmentation with RGB image classification using ViT, ResNet, and EfficientNet. User-defined prompts and reference image embeddings enhance segmentation and classification flexibility. Point-based prompting achieved the best IoU, while embedding-based majority voting supported modular adaptability. The results underscore the importance of dataset diversity over preprocessing and highlight a scalable approach for efficient annotations and broader agricultural applications.

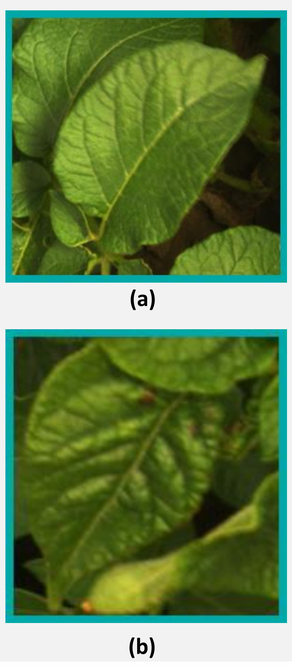


Figure 1. (a) Example of Healthy leaf. (b) Example of PVY-infected leaf.

Materials and Methods

- Dataset: 32 images with 2291 leaf samples (1176 healthy, 1115 PVY-infected)
- Classifier trained on individual leaves
- Segmentation via SAM2 to extract ROIs and Classification of segmented ROIs
- Extracted feature embeddings from classifier for reference image guided classification

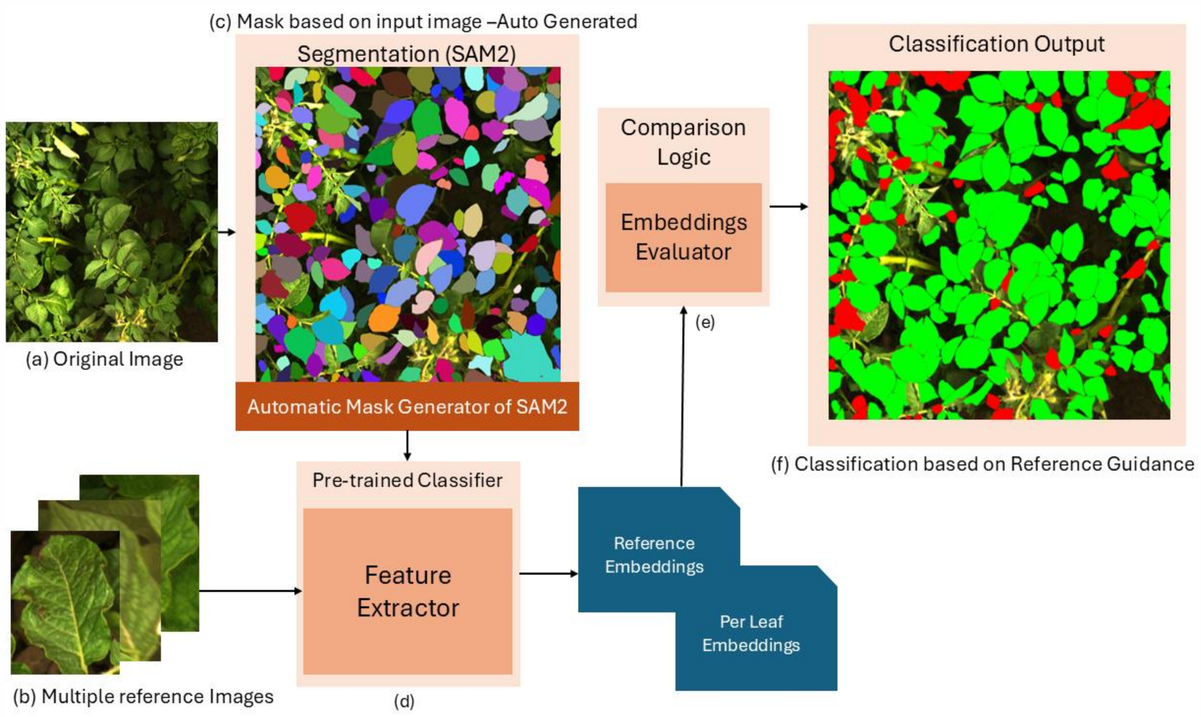


Figure 2. Reference Image-Guided Workflow: A workflow illustrating the process from data input to the final prediction, using a reference image as a guide for classification. (a) Original image used for segmentation and classification. (b) Multiple reference images provided as input. (c) Masks generated by the Automatic Mask Generator of SAM2. (d) Feature extraction logic using a pre-trained classifier. (e) Distance comparison of feature embeddings. (f) Classification results based on reference images, with green masks indicating healthy leaves and red masks representing PVY-infected leaves.

Experiments and Results

- **Experiment 1**: Classifier models were tested with varying blur thresholds.

Table 1. Performance Metrics: Comparison of classifier models across blur levels, showing F1 Score (%) and blur leaves removed.

Standard Deviation of Blur	ViT-B/16 F1 score (in %)	ResNet-18 F1 score (in %)	EfficientNet-B0 F1 score (in %)	Number of Blurry leaves removed
-2.0	**75.12**	66.77	69.83	0
-1.0	72.94	68.60	66.64	167
-0.75	69.60	70.83	70.97	402
-0.5	68.86	64.51	69.32	660
-0.25	70.12	65.35	62.34	866



Figure 3. Example of blurry leaf.

- **Experiment 2**: Classifier models were tested under different embedding comparison techniques.

Table 2. Performance Metrics: Comparison of classifier models across different embedding techniques, showing F1 Score (%).

Comparison Techniques	ViT-B/16 F1 score (in %)	ResNet-18 F1 score (in %)	EfficientNet-B0 F1 score (in %)
Single Feature Embedding	61.50	64.35	62.64
Average Feature Embedding	74.16	70.26	76.12
Majority Voting	77.25	**78.03**	**78.03**

- **Experiment 3**: Prompting techniques were tested for segmentation accuracy using IoU.

Table 3. Comparison of average IoU scores for Point, Box, and Edge prompting techniques.

Prompt Technique	Average IoU (in %)
Point	**86.00**
Edge	31.87
Box	76.00



Figure 4. Prompting Workflow: (a) Original image for segmentation and classification. (b) Point prompt. (c) Box prompt. (d) Masks generated by SAM2. (e) Classification overlay from a pre-trained model based on user prompts.



(a) Point Prompt **(b) Edge Point Prompt** **(c) Box Prompt**

Figure 5. Different Prompt Techniques: Black indicates the user-defined prompts (point or box). The green box represents the ground truth, while the red box denotes the segmentation prediction.

Conclusions

- Larger, diverse datasets outperform preprocessing like removing blurry samples for classification.
- ViT-B/16 excels in handling image quality variations, surpassing ResNet18 and EfficientNet-B0.
- Reference image-based approach empowers users to guide classification with feature embeddings, avoiding reliance on black-box inputs.
- Point-based prompting works best for dense agricultural setups.
- Instance segmentation with masks and boxes simplifies annotation, helping experts review and feed data into models easily.

Acknowledgements

- This project is financially supported by the European Union within the JTF-call project DigiAgro3.

Medegefinancierd door de Europese Unie

References

- [1] Christine L Carroll, Colin A Carter, Rachael E Goodhue, and C.-Y. Cynthia Lin Lawell. Crop disease and agricultural productivity. Working Paper 23513, National Bureau of Economic Research, June 2017.
- [2] G. Polder, P. M. Blok, H. A. C. De Villiers, J. M. Van der Wolf, and J. Kamp. Potato virus y detection in seed potatoes using deep learning on hyperspectral images. Frontiers in plant science, 10:209, 2019.