HANIF
Journal of Information Systems
https://journal.ilmubersama.com/index.php/hanif

Click here and write your Article Category

Computer Vision-Based Lettuce and Plantweed Segmentation Using YOLO and Segment Anything for Precision Agriculture

Akhmad Jayadi[1], Adi Ahmad Fauzi[2], Muhammad Ikhsan[3], Jaka Persada Sembiring[4], Ahmad Rofi'i[5]

[1,5] *Informatics Management, Politeknik Negeri Lampung, Bandar Lampung, Indonesia*
[2] *Informatics, Universitas Muhammadiyah Lampung, Bandar Lampung, Indonesia*
[3] *Computer Science, Universitas Lampung, Bandar Lampung, Indonesia*
[4] *Electrical Engineering, Teknokrat University, Bandar Lampung, Lampung, Indonesia*

ARTICLE INFORMATION

Received:February 00, 00
Revised:March 00, 00
Available Online:April 00, 00

KEYWORDS

YOLO11s; Segment Anything Model; Image Segmentation; Computer Vision; Precision Agriculture

CORRESPONDENCE

E-mail: akhmad.jayadi@polinela.ac.id

A B S T R A C T

The presence of weeds in lettuce cultivation areas can reduce plant productivity because they compete for nutrients, water, and light. Manual weed identification requires considerable time and effort, so an automated system based on *computer vision is needed* . This study proposes the integration of the YOLO11s model and the Segment Anything Model (SAM) to detect and segment lettuce plants and weeds in agricultural environments. The dataset used consists of 741 training images **,** 212 validation images , and 106 testing images with two object classes, namely *Lettuce* and *Plantweed* . The YOLO11s model was trained for 100 epochs using an image size of 640 × 640 pixels and *a batch size of 106* . 16. The training results show that the model obtained an mAP@50 value of 87.8%. And mAP@50–95 was 80.0% , indicating good object detection capability. Furthermore, the *bounding box coordinates* of the detection results were used as *prompts* in the Segment Anything Model to generate *mask -shaped segmentations* that follow the object contours more precisely. The experimental results show that the integration of YOLO11s and SAM is able to produce more detailed object representations compared to detection using *bounding boxes* alone. This approach has the potential to support various precision agriculture applications, such as plant morphology analysis, leaf area estimation, and the development of automated weed control systems .

INTRODUCTION

Precision agriculture [1] [2] is a modern approach that utilizes digital technology to increase the efficiency of crop cultivation through more accurate land management. One of the main challenges in cultivating horticultural crops [3] , especially lettuce (Lactuca sativa), is the presence of weeds [4] growing around the plants. Weeds compete with the main crop [5] for nutrients, water, and light, thus reducing plant growth and productivity if not properly controlled. The process of weed identification in the field is still largely done manually, so it takes a long time, requires a large workforce, and is prone to observation errors, especially in areas with high plant populations.

The development of computer vision [6] and deep learning [7] technology opens up opportunities for automatic plant identification using digital images. One widely used algorithm is You Only Look Once (YOLO) [8] , which is known for its rapid object detection capabilities with a high level of accuracy [9] . YOLO has been widely applied in various fields, including agriculture [10] , because it is able to detect various objects in real-time using a single inference process. In agricultural applications, YOLO is used to detect plants, fruits, plant diseases, and weeds based on the position of objects represented in the form of bounding boxes [11] . This approach is very effective for determining the location of objects, but still has limitations in describing the actual shape of the plant because the area inside the bounding box still includes part of the background.

https://doi.org/10.12345/xxxxx

To obtain a more detailed representation of objects, image segmentation techniques are a more appropriate alternative than object detection alone. Segmentation [12] allows each pixel in the image to be classified as part of an object or background so that plant contours can be obtained more accurately. This information is very important in various precision agriculture applications, such as calculating leaf area, analyzing plant growth, estimating weed cover, and developing selective herbicide spraying systems. Therefore, the combination of object detection and image segmentation is an approach that is increasingly being developed in computer vision-based research. One of the segmentation models that is currently widely used is the Segment Anything Model (SAM) [13] [14] , which is a foundation model developed to produce object segmentation on various types of images without requiring special retraining. SAM is able to produce object masks by utilizing various types of input, such as points, bounding boxes [15], and initial masks. The integration between YOLO as an object detector and SAM as a segmentation model allows the segmentation process to be carried out in a more targeted manner, because the bounding box from the detection is used as input for SAM to produce a mask that follows the shape of the object precisely.

Several previous studies have utilized YOLO to detect crops and weeds in agricultural environments. Other studies have also shown that segmentation models can improve the quality of object representation compared to bounding box-based detection approaches. However, most studies still focus on the object detection or segmentation stages separately. Research that integrates detection using YOLO with SAM-based segmentation [16]for lettuce and weed identification in real agricultural environments is still relatively limited, especially in field datasets that have variations in lighting, plant size, and complex background conditions. Based on these problems, this study proposes a computer vision-based lettuce and weed segmentation system by combining the YOLO11s model [17]as an object detector and the Segment Anything Model as a segmentation model. The YOLO11s model is used to detect the location of lettuce and weed plants in field images, while the detection results in the form of bounding boxes are used as input for SAM to produce segmentation in the form of masks that follow the contours of objects.

METHOD

This study uses a computer vision approach that integrates the YOLO11s object detection model with the Segment Anything Model 2 (SAM 2) to segment lettuce and weeds in agricultural imagery. In general, the research stages consist of dataset preparation, YOLO11s model training, object detection, segmentation using SAM, Region of Interest (ROI) extraction, and model performance evaluation. The research flowchart is shown in Figure 1.

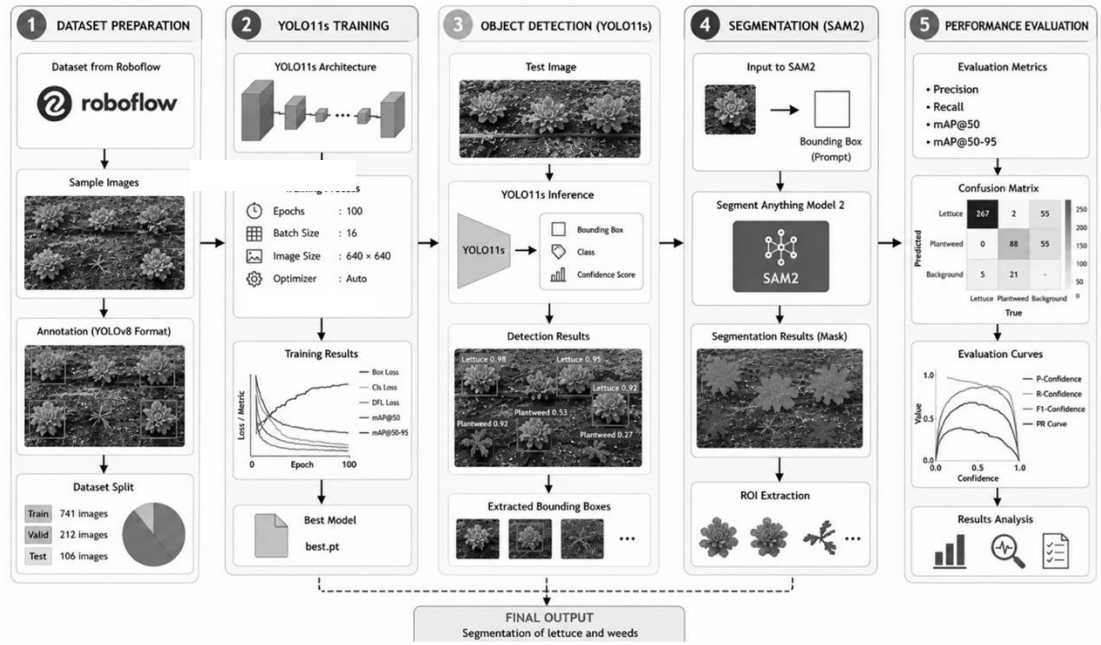

Figure 1. Research Flowchart

The first stage was collecting a dataset obtained from Roboflow in YOLOv8 format. The dataset consisted of images of lettuce plants and weeds, each with annotated bounding boxes. The dataset was then divided into training, validation, and testing data for the model development process.

In the next stage, the YOLO11s model was trained using the training dataset until the best weights were obtained. The trained model was then used to detect the position of lettuce and weed objects in the test image. The resulting bounding box coordinates were then used as input (prompts) for the Segment Anything Model to generate mask-shaped segmentations that more accurately follow the object's contours.

The segmentation results were then used to obtain a Region of Interest (ROI) for each object, allowing the plant area to be separated from the background. The final stage was model performance evaluation using Precision, Recall, mAP@50, and mAP@50–95 metrics, as well as visual analysis of the detection and segmentation results.

Dataset

The dataset used in this study was obtained from the Roboflow platform and annotated using the YOLOv8 format. The dataset consists of two object classes: Lettuce as the main crop and Plantweed as a weed. All annotations were performed using bounding boxes indicating the location of each object in the image.

The dataset was divided into three groups: 741 training images, 212 validation images, and 106 testing images. All images had an input size of 640 × 640 pixels, which was used consistently throughout the training and inference processes. Examples of dataset images and their annotations are shown in Figure 2.

```
1 0.18658203125 0.9662890625 0.05197265625 0.067421875
1 0.23931640625 0.648515625 0.111435546875 0.10864583333333333
1 0.230576171875 0.5486067708333333 0.0546875 0.06134114583333333
0 0.48681640625 0.5143229166666666 0.2783203125 0.3802083333333333
0 0.91015625 0.49609375 0.173828125 0.3463541666666667
```

Figure 2. Example of an image and annotation

Dividing the dataset into three parts, namely train, test, and valid, aims to reduce the risk of overfitting while obtaining a more objective model evaluation of data that has never been used before.

YOLO11s Model Training

Object detection in this study was conducted using the YOLO11s model, a lightweight variant of the YOLO11 family. The model was chosen because it balances accuracy and inference speed, making it suitable for precision agriculture applications that require rapid processing.

Model training was performed using the Ultralytics YOLO framework with parameters as shown in Table 1.

Table 1. YOLO11s Training Parameters

Parameter	**Mark**
Model	YOLO11s
Epoch	100
Batch Size	16
Image Size	640 × 640

 https://doi.org/10.12345/xxxxx

Optimizer	Auto
Number of Classes	2

During the training process, the model learns object positions based on bounding box annotations. At each epoch, an evaluation is performed using validation data to obtain the best model (best.pt) for use in the inference stage.

Object Detection Using YOLO11s

YOLO (You Only Look Once) is a deep learning algorithm widely used for real-time object detection. Unlike two-stage detectors, YOLO performs feature extraction, object location prediction (bounding box), and object classification in a single network (single-stage detector). This approach achieves high inference speed while maintaining a good level of accuracy, making it widely applied in various fields such as autonomous vehicles, robotics, precision agriculture, and visual surveillance. This study used YOLO11s, a small variant of the YOLO11 family developed by Ultralytics. This model was chosen because it offers a good balance between detection accuracy and computational requirements, making it suitable for precision agriculture applications that require rapid image processing.

The YOLO11 architecture consists of several key modules that play a role in improving feature extraction quality and computational efficiency. The Spatial Pyramid Pooling Fast (SPPF) module is used to expand the receptive field so that the model can capture object information at various scales. Furthermore, the C2PSA module integrates an attention mechanism to enhance the representation of important features while suppressing less relevant information. Furthermore, the C3K2 module improves the efficiency of the feature extraction process with lower computational complexity compared to conventional convolution blocks. Through the combination of these three modules, YOLO11 is capable of supporting various computer vision tasks such as object detection, instance segmentation, image classification, pose estimation, oriented object detection, and object tracking. In this study, the object detection capability was utilized to obtain the locations of lettuce plants and weeds, which were then used as input for the segmentation process using the Segment Anything Model (SAM2).

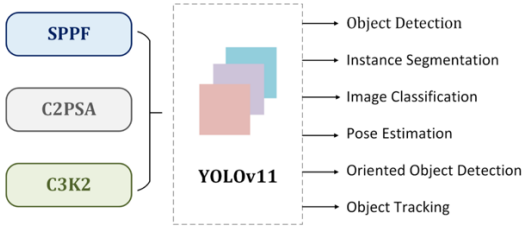

Figure 3. Key architectural modules in YOLO11 [17]

Segmentation Using the Segment Anything Model

The segmentation stage is performed using the Segment Anything Model (SAM). Unlike YOLO, which only produces a bounding box, SAM produces a pixel mask that follows the object's shape in greater detail. In this study, each bounding box detected by YOLO is used as a prompt for SAM to generate object segmentation. This approach allows the segmentation process to focus only on the detected areas, thus reducing the possibility of segmenting irrelevant objects.

The segmentation mask is then used to obtain a Region of Interest (ROI) for each object, allowing for a more accurate representation of the plant's shape compared to using only a bounding box. An example of segmentation results using SAM is shown in Figure 4.

Figure 4. Example Image [13]

Model Evaluation

Model performance evaluation is performed based on metrics commonly used in object detection tasks, namely Precision, Recall, mean Average Precision at IoU 0.5 (mAP@50), and mean Average Precision in the IoU range 0.5–0.95 (mAP@50–95). Precision is used to measure the proportion of correct positive predictions against all positive predictions with the formula which can be expressed as

$$P = \frac{TP}{TP + FP} \tag{1}$$

while Recall shows the model's ability to find all objects that are actually present in the image with the formula it can be expressed as

$$R = \frac{TP}{TP + FN} \tag{2}$$

TP is true positive, which indicates the number of positive classes that are actually predicted as positive; FP is false positive, which indicates the number of negative classes that are actually predicted as positive; FN is false negative, which indicates the number of positive classes that are actually predicted as negative [18].

The mAP@50 value is used to evaluate the detection accuracy at an Intersection over Union (IoU) threshold of 0.5, while mAP@50–95 provides a more comprehensive evaluation because it calculates the average model performance at various IoU values, with the mAP formula being expressed as

$$AP_i = \int_0^1 P_i(R)dR \tag{3}$$

$$mAP = \frac{\sum_1^N AP_i}{N} \tag{4}$$

In addition to quantitative evaluation, this study also conducted a visual analysis of the YOLO11s detection results and the segmentation results using SAM. This analysis aimed to demonstrate the ability of both models to more accurately represent plant objects in agricultural environments with varying plant shapes, object sizes, and background conditions.

RESULTS AND DISCUSSION

This section comprehensively reviews the implementation of a lettuce and plantweed segmentation system using a hybrid approach based on YOLO (You Only Look Once) and SAM2 (Segment Anything Model 2). Experiments were conducted on a Google Colab environment with NVIDIA Tesla T4 GPU hardware acceleration.

YOLO11s Model Training Results

The YOLO11s model was trained using a dataset consisting of 741 training images, 212 validation images, and 106 test images. The training process was carried out for 100 epochs with an image size of 640 × 640 pixels, a batch size of 16, and an auto-optimizer. The entire training process produced the best model (best.pt), which was then used in the inference stage.

The development of training loss and validation loss values during the training process is shown in Figure 5. Based on the graph, the box loss, classification loss, and distribution focal loss (DFL) values gradually decreased until they approached convergence at the end of training. The decrease in loss values indicates that the model successfully learned the characteristics of lettuce and weed objects effectively without showing any significant indication of overfitting.

In addition to the loss reduction, the Precision, Recall, mAP@50, and mAP@50–95 values also experienced a steady increase during the training process. The mAP@50 value reached around 87.8%, while the mAP@50–95 value approached 80% at the end of training. These results indicate that the model has good capabilities in detecting both object classes at various Intersection over Union (IoU) levels.

Overall, the training curve shows that the model has reached a convergent state before the training ends so that the best model is able to provide a balance between generalization ability and detection accuracy.

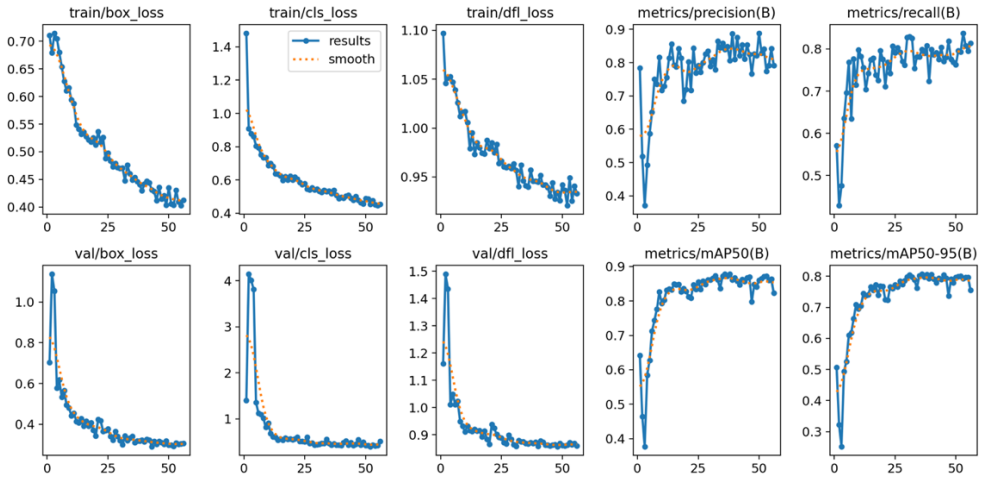

Figure 5. YOLO11s training result curve

Model Performance Evaluation

Model evaluation was performed using Precision, Recall, Mean Average Precision (mAP), and Confusion Matrix metrics. Based on the training results, the model achieved a Precision value of around 0.84, Recall of around 0.80, mAP@50 of 0.878, and mAP@50–95 of 0.800. These values indicate that the model is capable of detecting most objects with a high level of accuracy.

The Precision–Recall curve in Figure 6 shows that the Lettuce class performed very well with an Average Precision (AP) value of 0.978, while the Plantweed class achieved an AP of 0.779. This difference indicates that lettuce objects are easier to recognize than weeds due to their larger size, more consistent leaf shape, and a larger number of samples in the dataset.

The F1-Confidence curve shows that the highest F1 value is achieved at a confidence threshold of around 0.397, with an F1 value of around 0.80. This indicates that using a confidence value around 0.40 provides the best balance between Precision and Recall.

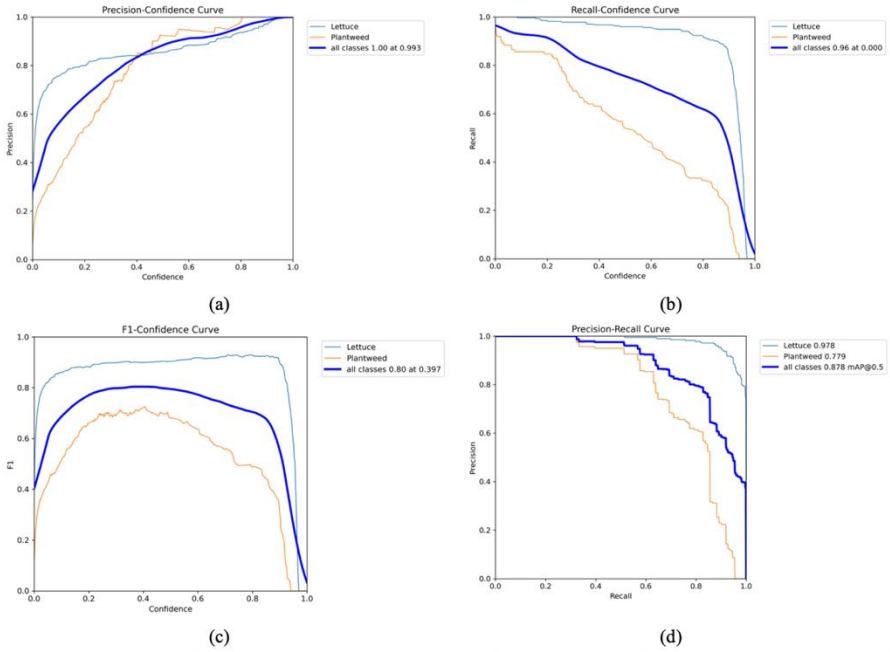

Figure 6. Evaluation curves of (a) Precision, (b) Recall, (c) F1, and (d) Precision–Recall

Confusion Matrix Analysis

The Confusion Matrix of the test results is shown in Figure 7. The matrix describes the relationship between the actual class and the model's predicted results.

Based on the Confusion Matrix, the majority of Lettuce objects were correctly classified. This indicates that the model consistently recognizes the visual characteristics of lettuce plants. Conversely, several classification errors were still found in the Plantweed class, particularly for small weeds or those with leaves similar to lettuce.

Furthermore, there are still a number of objects that were not detected and are therefore categorized as background. This condition generally occurs with objects that are very small, partially obscured by other plants (occlusion), or have low contrast against the background ground.

In general, the results of the Confusion Matrix show that the model has good classification capabilities, although there is still room for performance improvement, especially in the weed class which has more varied shapes and sizes than lettuce plants.

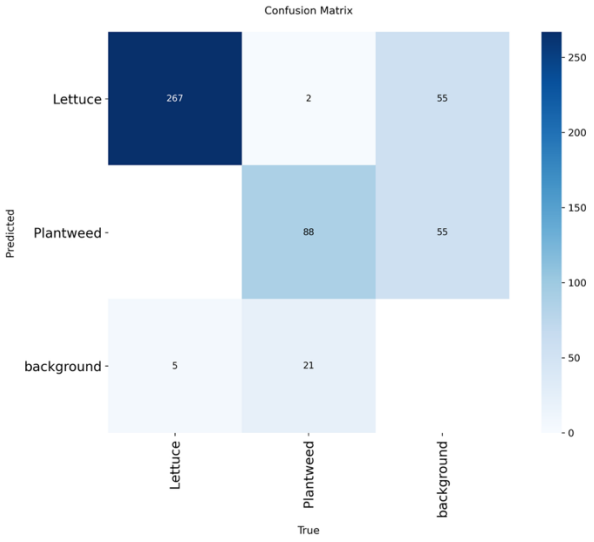

Figure 7. Confusion Matrix

 https://doi.org/10.12345/xxxxx

Detection Results Using YOLO11s

The next stage is the inference process using the YOLO11s model on the test image. The detection results are shown in Figure 8. The model successfully detected several lettuce plants and surrounding weeds. Each object was represented using a bounding box complete with class labels and confidence values. For most lettuce objects, the model produced confidence values above 0.90, indicating a high level of prediction confidence. However, some weed objects had relatively low confidence values, even below 0.40. This is likely due to the small size of the weeds, the smaller sample size compared to the lettuce class, and the similarity in color between the weed leaves and the main crop. Furthermore, the presence of shadows and variations in lighting in the field environment also affected detection quality. However, the detection results show that YOLO11s is able to identify the location of objects well so that the bounding box coordinates can be used as a prompt for the segmentation process using the Segment Anything Model.

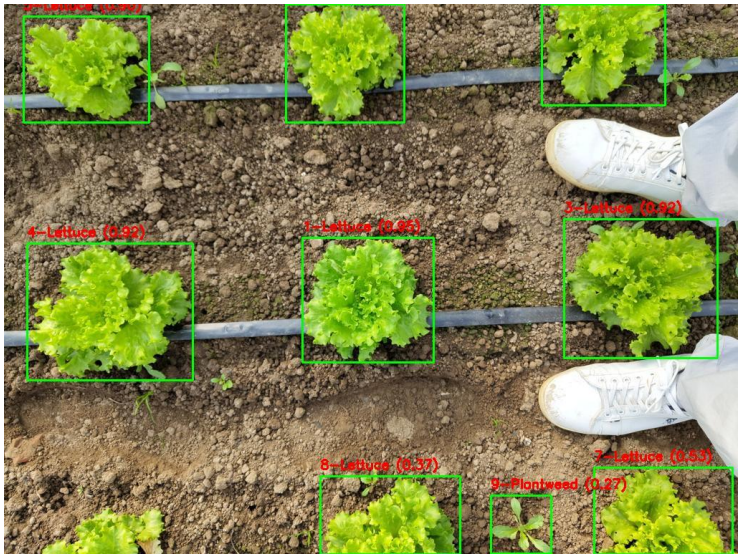

Figure 8. YOLO11s detection results

Segmentation Results Using the Segment Anything Model

The final stage of the research was object segmentation using the Segment Anything Model (SAM). Unlike YOLO, which only produces bounding boxes, SAM produces masks that follow the contours of the plants in greater detail. In this study, the bounding box coordinates from YOLO detection were used as prompts, allowing SAM to segment only the detected object area. This approach yields more precise segmentation than segmentation without object location information.

The segmentation results in Figure 9 show that the mask successfully follows the shape of the lettuce leaf in detail, including the wavy edges. For the weed object, SAM also managed to separate it from the background soil quite well, despite its relatively small size. The integration of YOLO11s and SAM offers advantages over object detection alone. YOLO rapidly determines object location, while SAM refines object representation through pixel-based segmentation. This representation enables the extraction of advanced information such as leaf area, plant morphology, and weed coverage, as well as the development of selective herbicide spraying systems for precision agriculture applications.

Overall, the combination of YOLO11s and the Segment Anything Model demonstrates a complementary approach. YOLO provides high-efficiency detection capabilities, while SAM improves object representation, enabling the system to generate richer visual information to support crop analysis in agricultural environments.

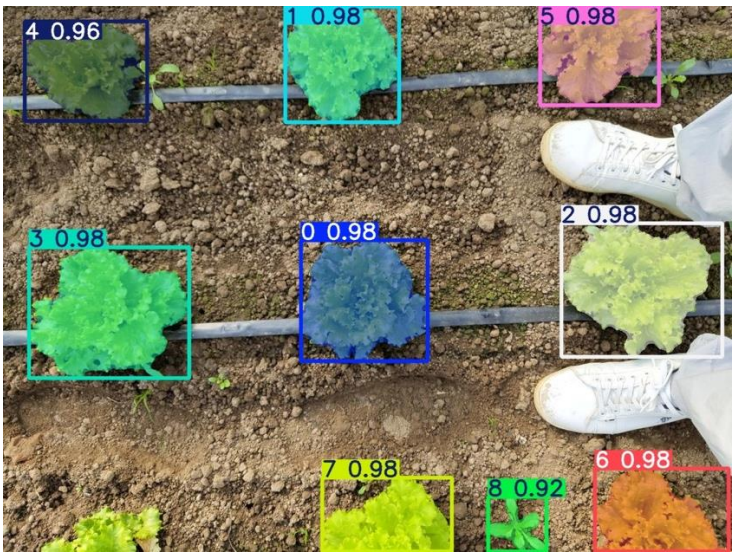

Figure 9. Segmentation results using SAM

CONCLUSION

This study successfully developed a computer vision-based lettuce and weed segmentation system by integrating the YOLO11s model and the Segment Anything Model (SAM). The YOLO11s model was used to detect the location of plant and weed objects in agricultural images, while the detection results in the form of bounding boxes were used as prompts for SAM to produce mask-shaped segmentations that follow the object contours more precisely. Based on the training results using 741 training data, 212 validation data, and 106 test data, the YOLO11s model showed good detection performance with a mAP@50 value of 87.8% and a mAP@50–95 of 80.0%. The evaluation results also showed that the model was able to effectively distinguish lettuce and weed objects although there were still some detection errors in small weed objects or those that overlapped with the main crop.

The integration of YOLO11s and the Segment Anything Model successfully improves object representation through pixel-based segmentation. Unlike detection using bounding boxes alone, the segmentation results are able to follow leaf shapes more accurately, providing more complete visual information about plant characteristics. This representation has the potential to be used for various precision agriculture applications, such as plant morphology analysis, leaf area estimation, weed growth identification, and the development of automated weed control systems. In future research, the system can be developed using end-to-end segmentation models, such as YOLO-Seg or Mask R-CNN, and then compared with the YOLO-SAM approach to evaluate differences in accuracy, inference speed, and computational requirements. Furthermore, the use of more diverse datasets with varying lighting conditions, plant growth stages, and more weed types is also expected to improve the model's generalizability in real agricultural environments.

REFERENCES

[1] Iwan Patria, “Transformasi Pertanian Presisi Berbasis Digital di Indonesia: Tinjauan Sistematis Peluang dan Tantangan,” *INSOLOGI J. Sains dan Teknol.*, vol. 4, no. 6, pp. 1462–1477, Dec. 2025, doi: 10.55123/insologi.v4i6.6319.

[2] R. Rosidi, “Pengembangan Pertanian Presisi Berbasis Teknologi Digital di Indonesia Perspektif Al-Qur’an,” *Alashriyyah*, vol. 11, no. 2, pp. 345–364, Oct. 2025, doi: 10.53038/alashriyyah.v11i2.282.

[3] F. R. Aminda, H. Anggrasari, and A. K. Sari, “Study of Horticultural Crop Leading Commodities Development in Banjarnegara Regency, Central Java Province,” *Agritech J. Fak. Pertan. Univ. Muhammadiyah Purwokerto*, vol. 25, no. 2, p. 163, Feb. 2024, doi: 10.30595/agritech.v25i2.19566.

[4] R. Sabtu and S. Kisman, “IDENTIFIKASI GULMA PADA LAHAN PERCONTOHAN PKK KELURAHAN MALIARO KOTA TERNATE,” *SAINTIFIK@ J. Pendidik. MIPA*, vol. 9, no. 2, pp. 1–11, Nov. 2024, doi: 10.33387/saintifik.v9i2.8958.

[5] P. L. Tarigan, A. Fitrianti, A. W. Wardhana, V. Gabrielle, S. I. Andisha, and N. Ayni, “Analisis Vegetasi Gulma Tanaman Mawar pada Lahan Dataran Tinggi dan Rendah,” *J-Plantasimbiosa*, vol. 7, no. 1, pp. 38–44, May 2025, doi: 10.25181/jplantasimbiosa.v7i1.3880.

[6] D. R. Tobergte and S. Curtis, *Algorithms for image prcessing and computer vision*, vol. 53, no. 9. 2013. doi:

10.1017/CBO9781107415324.004.

[7] D. S. Alfan and I. Kumalasari, "Implementasi Model Deep Learning MobileNetV2 untuk Klasifikasi Citra Melanoma Berbasis Web," vol. 7, no. 3, pp. 670–680, 2026, doi: 10.47065/josh.v7i3.8848.

[8] Reni Triyaningsih, Pradita Eko Prasetyo Utomo, and Benedika Ferdian Hutabarat, "Application of You Only Look Once (YOLO) Method for Sign Language Identification," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 14, no. 4, pp. 254–262, Nov. 2025, doi: 10.22146/jnteti.v14i4.21931.

[9] I. A. Zulkarnain and K. Kusrini, "Optimasi Yolov11 Melalui Hyperparameter Tuning dan Data Augmentasi untuk Meningkatkan Akurasi Deteksi Kendaraan pada Kondisi Malam Hari: Yolov11 Optimization Through Hyperparameter Tuning and Data Augmentation to Improve Vehicle Detection Accuracy at Night," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 4, pp. 294–1303, 2025, doi: 10.57152/malcom.v5i4.2250.

[10] K. Krisdianto, Elta Sonalitha, and Yandhika Surya Akbar Gumilang, "Deteksi penyakit padi menggunakan YOLO," *Uranus J. Ilm. Tek. Elektro, Sains dan Inform.*, vol. 2, no. 3, pp. 125–134, Jul. 2024, doi: 10.61132/uranus.v2i3.259.

[11] H. Herdianto, H. Hafni, D. Nasution, and S. Ramadhan, "Implementasi Metode Yolo pada Deteksi Objek Manusia," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 8, no. 2, pp. 234–240, Oct. 2024, doi: 10.46880/jmika.Vol8No2.pp234-240.

[12] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3059968.

[13] A. Kirillov *et al.*, "Segment Anything," 2023.

[14] L. Zhang, X. Deng, and Y. Lu, "Segment Anything Model (SAM) for Medical Image Segmentation: A Preliminary Review," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Dec. 2023, pp. 4187–4194. doi: 10.1109/BIBM58861.2023.10386032.

[15] M. Luo, T. Zhang, S. Wei, and S. Ji, "SAM-RSIS: Progressively Adapting SAM With Box Prompting to Remote Sensing Image Instance Segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024, doi: 10.1109/TGRS.2024.3460085.

[16] P. Ghose, A. Bashir, Y. Wang, C. Bua, and A. Zahid, "YOLO-SAM AgriScan: A Unified Framework for Ripe Strawberry Detection and Segmentation with Few-Shot and Zero-Shot Learning," *Sensors*, vol. 25, no. 24, p. 7678, Dec. 2025, doi: 10.3390/s25247678.

[17] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," vol. 2024, pp. 1–9, 2024, [Online]. Available: http://arxiv.org/abs/2410.17725

[18] Z. Wang *et al.*, "PC-YOLO11s: A Lightweight and Effective Feature Extraction Method for Small Target Image Detection," *Sensors*, vol. 25, no. 2, p. 348, Jan. 2025, doi: 10.3390/s25020348.