A Segment Anything Model Approach for Rice Seedlings Detection Based on UAV Images

Hassan Rezvan [1], Mohammad Javad Valadan Zoej [1], Fahimeh Youssefi [1,2]

[1] Department of Photogrammetry and Remote Sensing, K. N. Toosi University of Technology 19967-15433, Tehran, Iran – (h.rezvan@email.kntu.ac.ir, valadanzouj@kntu.ac.ir, youssefi@email.kntu.ac.ir)
[2] Institute of Artificial Intelligence, USX, Shaoxing University, 508 West Huancheng Road, Yuecheng District, Shaoxing, Zhejiang Province, Postal Code 312000, China – youssefi@usx.edu.cn

Keywords: Remote sensing, Smart farming, Rice seedling, Transfer learning, Segment anything model, UAV.

Abstract

Accurate estimation of regional rice yields is crucial for food security and efficient agricultural management. In this regard, the use of Unmanned Aerial Vehicles (UAVs) that have revolutionized crop monitoring by providing high-resolution images for precision agriculture, is beneficial. This study explores the potential of Segment Anything Model (SAM) for detecting rice seedlings, focusing on determining the optimal approach and prompt for this task. We examined three SAM scenarios: automatic mask generation, bounding box prompt, and point prompt. Our evaluation criteria included processing time, visual interpretation, and accuracy indexes. The results demonstrated the effectiveness of SAM in rice seedling detection, highlighting the importance of selecting the appropriate prompt for specific agricultural applications. Our findings reveal that the point prompt method emerges as the preferred choice for rice seedling detection, offering superior accuracy and reliability. Specifically, it achieved mIoU and mDice scores of 94.57 % and 0.97, respectively. While the bounding box approach showed promise, despite slightly lower precision, it may still be suitable depending on application-specific requirements. Conversely, the automatic mask generation scenario proved unsuitable for this task due to its low accuracy and inability to effectively detect rice seedlings. The outcomes of this study serve as a baseline for evaluating SAM prompts, guiding future improvements and refinements to enhance its performance in real-world agricultural applications.

1. Introduction

Smart farming is an indispensable solution to the statistics provided by the World Resources Institute (WRI), according to which 10 billion people will need food by 2050 (Islam et al., 2021). Smart farming is a system that integrates cutting-edge technologies with traditional farming methods to enhance both the quality and quantity of agricultural production, while significantly reducing inputs (Islam et al., 2021; Hashim et al., 2023). Utilizing smart farming technologies will undoubtedly help farmers with a variety of tasks to boost rice production (Hashim et al., 2023). Therefore, it plays a crucial role in optimizing agricultural management by providing insights and recommendations for more efficient and effective production, helping to address challenges in agricultural systems. The primary aim of smart farming is to boost productivity, increase yields, and improve profitability, all while minimizing the environmental impact through techniques such as efficient irrigation and the precise application of pesticides and fertilizers (Islam et al., 2021; Tseng et al., 2022).

Rice is the primary and the second world’s largest, staple food that feeds more than half of the global human population with more than 90% of rice production being in Aisa (Wu et al., 2019; Xiao et al., 2021; Anuar et al., 2022). Sufficient yielding of grain crops has been considered in many countries as one of the most important issues to maintain food security (Yang et al., 2021). Therefore, methods for estimating rice yield have received significant attention (Wu et al., 2019).

Moreover, the number of rice seedlings per unit area is an important agronomic component which influence the quantity and quality of production (Wu et al., 2019). It is not only closely associated with yield, but also plays an important role in the determination of survival rate. Additionally, it is necessary for farmers to inspect the paddy fields to search for defective paddy seedlings and replace or replant them manually (Anuar et al., 2022). To achieve this, accurate detecting and counting of rice seedlings is crucial for predicting rice yields and is a fundamental step in field phenotyping. Currently, manual human effort remains the primary method for counting rice seedlings in the field. This manual approach is time-consuming and labor-intensive, especially for researchers conducting large-scale measurements. Consequently, there is an urgent need to develop a fast, non-destructive, and reliable technique that can accurately detect seedlings and count them in the field without causing damage to the plants (Wu et al., 2019).

Remote sensing has become a popular technique for obtaining crop information in plant phenotyping, due to its ability to capture multi-temporal imaging and data acquisition in a large area (Wu et al., 2019). However, satellites that are limited by spatial resolution, cannot provide highly detailed data for a precise agriculture monitoring (Yang et al., 2021). Thanks to the development of mechanical and electronic techniques, UAVs have been used in remote sensing fields and allows for observation of high-resolution spatial patterns for crop monitoring (Wu et al., 2019; Yang et al., 2021). Furthermore, UAVs have been incorporated into smart farming towards the aim of providing additional perspectives, such as, imagery analysis and agricultural surveillance. UAVs not only facilitate image analysis and processing of the agricultural area, but also, offer in-depth situation awareness by patrolling over an area of interest (Islam et al., 2021; Tseng et al., 2022).

Image-based techniques undoubtedly offer a practical, cost-effective, and efficient approach for crop counting. However, due to the complexity of field environments and factors like varying light conditions, soil reflectance, textures, and shapes, achieving accurate crop counting remains a significant challenge (Wu et al., 2019). Over the past decade, object detection based on machine learning and deep learning, has been paid enormously to improve counting crowds in terms of accuracy and speed (Wu et al., 2019; Yang et al., 2021; Anuar et al., 2022; Tseng et al., 2022). Although few studies have been conducted in crop-related object detection tasks (Wu et al., 2019).

There are several approaches for image segmentation manual or semi-automatic methods, which are time-consuming and labor-intensive. In recent years, deep learning models have been developed to automate the segmentation process. However, these models generally require large amounts of training data and are computationally expensive, making them impractical for many applications (Wu and Osco, 2023). The Segment Anything Model (SAM), introduced by Meta AI, enables segmentation of new and unseen objects using minimal prior training, thanks to its versatile prompt-based architecture (Kirillov et al., 2023). It achieves this through a zero-shot learning mechanism, having been trained on an extensive dataset. This allows SAM to generalize to a wide variety of object types and scenes beyond its training examples. As a result, practitioners can perform segmentation tasks without the need to create new annotated datasets or only by minimal user inputs such as points, bounding boxes, or text prompts (Kirillov et al., 2023; Osco et al., 2023; Wu & Osco, 2023). This is a great alternative for our situation, where we lack ground-truth data on rice seedlings.

The Vision Transformer (ViT) is a recent innovation in deep learning that has shown strong potential in image segmentation tasks. Unlike CNNs, which use convolutional operations, ViT utilizes self-attention mechanisms that enable it to capture long-range relationships and the overall context within images. This method has achieved competitive results across various computer vision applications, including remote sensing image segmentation, where it currently outperforms CNNs. ViT comes in several versions—ViT-H, ViT-L, and ViT-B—each with unique computational demands and architectural differences (Osco et al., 2023).

For this study, we utilized the SamGeo Python package (Osco et al., 2023; Wu and Osco, 2023) to operationalize the SAM model on UAV imagery, enabling flexible prompt-based segmentation of rice seedlings. This package enables the application of different SAM model prompts on various data sources, such as UAV and satellite imagery, within the Python programming environment. Additionally, by offering multiple ViT model options, it supports flexible processing and output generation in various formats. The development of this package has introduced a versatile and powerful tool for conducting image segmentation tasks in Python.

In this paper, we aimed to explore if zero-shot approach for SAM is suitable for rice seedling detection from UAV images in the first place, and then which of its models and prompts gives the best and most optimal result for this application to provide a basis for predicting the amount of yields.

The rest of the paper is structured as follows. In Section 2, we give an overview of previous researches related in this area. In Section 3 our materials and methods are presented. Section 4 shows the results and discusses and compares them. And finally, in Section 5, conclusions are outlined.

2. Related works

As stated earlier, rice seedling detection has become increasingly important in modern agriculture due to its critical role in optimizing crop yields and reducing unnecessary resource allocation. This operation is not only limited to rice cultivation; similar techniques are employed for detecting seedlings in various crops and products. In this section, we aim to comprehensively examine the current state-of-the-art in rice seedling detection, exploring various methodologies, their performance metrics, and future directions for research in this field.

In 2019, Wu et al. presented a novel method for automatically counting rice seedlings from UAV images using deep learning techniques. The authors developed a combined network architecture consisting of two fully convolutional neural networks to address the challenges of accurately detecting and counting rice seedlings in real-world agricultural settings. The proposed method uses a Basic Network to estimate the distribution of rice seedlings and generate a density map, along with a Segmentation Network to separate rice seedling areas from non-rice seedling areas. To improve accuracy, the Combined Network was developed, which combines the outputs of both the Basic and Segmentation Networks to produce a final density map. Their method requires significant manual annotation time for training but offers advantages in speed and accuracy compared to alternative techniques like the Count Crops tool.

Yang et al. in 2021 presented a study on the use of UAVs in agricultural applications, specifically focusing on rice paddies. The authors provided a detailed annotated dataset aimed at enhancing deep learning practices in this field. For the classification task, the study modified a classical Convolutional Neural Network (CNN) architecture, specifically VGG-16. This modification was tailored to perform patch-based detection of rice seedlings, allowing for more precise localization and identification within the images. They used a cross-validation approach to validate their method and the results indicated that all divisions of the cross-validation dataset achieved an impressive accuracy of 99 %, demonstrating the effectiveness of the proposed methods. Another output of this paper was the provided dataset includes not only the training-validation dataset but also patch-based detection samples and an ortho-mosaic image of the field which is designed to facilitate further research and development in the application of deep learning to agricultural practices.

Anuar et al. in 2022 tried to enhance planting density by exploring the use of Deep Convolutional Neural Networks (DCNNs) for detecting defective paddy seedlings using aerial imagery. The authors experimented with different DCNN models for object detection, including EfficientDet-D1, SSD with MobileNetV2, SSD with ResNet50, and Faster R-CNN with ResNet50. The study successfully demonstrated that DCNN models could effectively detect defective paddy seedlings using aerial imagery, offering a scalable and efficient solution for improving paddy cultivation practices. The EfficientDet-D1 model was recommended for its balance between accuracy and processing speed.

In 2022, Tseng et al. detected rice seedlings in paddy fields using transfer learning from EfficientDet-D0 and Faster R-CNN and compared the results with histograms of oriented gradients-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-713-2025

based support vector machine (HOG-SVM) classification. CNN-based models outperformed the traditional HOG-SVM, especially in handling different environmental conditions and seedling sizes. They concluded that CNN-based models, especially Faster R-CNN, offer significant improvements in rice seedling detection. Transfer learning facilitates rapid model deployment with high detection accuracy.

3. Materials and methods

The flowchart of the proposed method is shown in Figure 1.

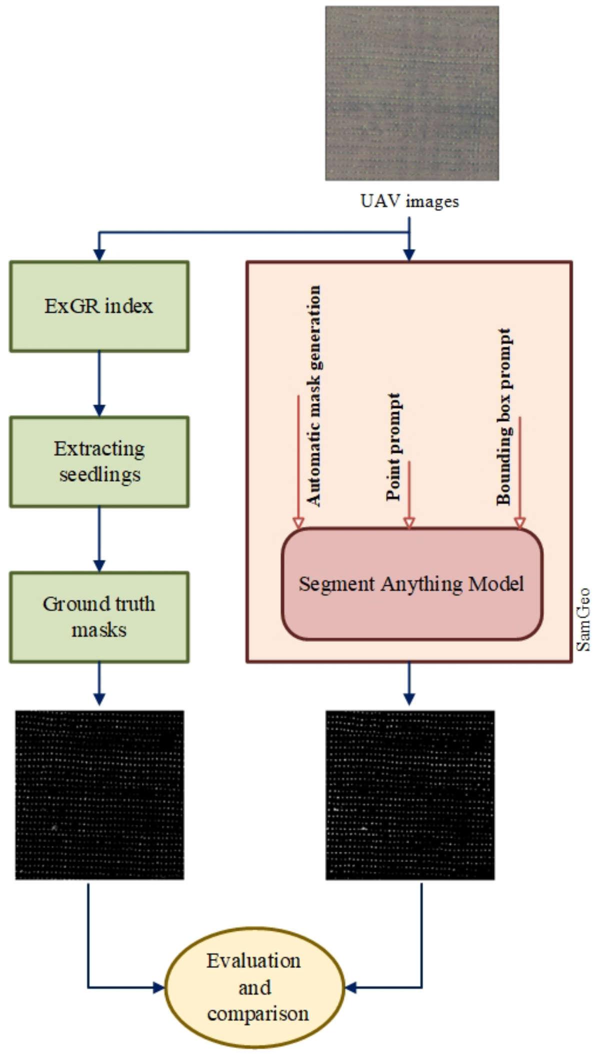

Figure 1. Flowchart of the proposed method.

In the proposed approach as shown in Figure 1, orthophotos of paddy rice fields are segmented using SamGeo library and by applying three distinct SAM prompts. Additionally, ground-truth masks for rice seedlings are created by applying the ExGR (Excess Green Ratio) index, differentiating their greenness from other regions. Finally, the performance of SAM prompts and their segmentation results are compared with the ground-truth masks and assessed using mIoU and mDice indices.

3.1 Dataset

The field images of rice seedlings were collected by UAVs equipped with cameras and downloaded from an open dataset (Yang et al., 2021). The dataset is a specialized collection of UAV images focused on rice paddies, designed to enhance deep learning and image segmentation applications in agriculture. It was acquired using a multi-rotor UAV that followed a planned scouting routine, ensuring consistent look-down perspectives essential for agricultural analysis. The dataset features a semi-automatic annotation method based on the ExGR index, which aids in generating training data specifically for rice seedlings. Also, it includes a training-validation dataset, patch-based detection samples, and an ortho-mosaic image of the field, providing a comprehensive resource for researchers.

For rice seedling detection, Yang et al. clipped 8 images from orthomosaic images with a region of 8 × 8 meters and a size of 1527 × 1527 pixels for each image. The object detection annotation of ground truth is also provided for the 8 images in XML files. A sample image of the dataset is shown in Figure 2.

Figure 2. An example of rice seedling detection dataset.

3.2 Methods

In this section, we describe how we evaluated the performance of SAM, for zero-shot approach in the context of UAV images to detect rice seedlings. As mentioned earlier, SAM is the first-of-its-kind, promptable segmentation model. Prompts allow to instruct the model on a desired output through text and interactive actions. Prompts could be provided to SAM in multiple ways: Points, Bounding Boxes, texts, and even base masks (Kirillov et al., 2023). Based on these prompts, three different scenarios have been considered and compared based on the criteria that will be explained below. Each of these scenarios is explained separately below:

3.2.1 Bounding box prompt: We utilized the SamGeo tool to process our geospatial analysis, specifically employing its bounding box feature. This functionality allowed us to leverage the coordinates of the object’s bounding boxes, which were extracted from XML annotation files (Figure 3). The tool then generated a list of geometric boundaries for our image data based on these coordinates. To optimize this process, we initialized a predictor instance, which facilitated the efficient handling of these boundaries. The predictor was used to segment and process the image. After being set up, the predictor analyzed each box and produced masks for the seedlings that had been segmented. This approach enabled the creation of individual instance segmentation masks for each bounding box's contents (Osco et al., 2023). These binary masks were subsequently merged and compiled into a single mosaic raster, resulting in a comprehensive visual representation of the segmented seedlings. This methodology allowed for a thorough examination of the image content while maintaining the ability to isolate and analyse specific areas of interest.

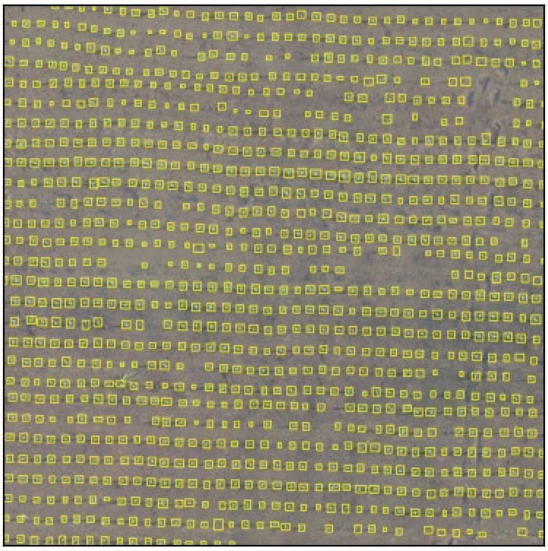

Figure 3. Rice seedlings annotated with bounding boxes in yellow as input for SAM bounding box prompt.

3.2.2 Point prompt: We adapted the bounding-box method to create a single-point feature prompt for our geospatial analysis. This new approach began with extracting the bounding box coordinates [x_min, y_min, x_max, y_max] from XML files, calculating the central point of each box to represent the seedlings (Figure 4). We once again utilized SamGeo for model prediction, but with a crucial difference: we set its automatic parameter to 'False' and applied the predictor to individual coordinates (Osco et al., 2023). Each point was iterated through, its features were predicted in certain cases, and the resulting mask was saved as a distinct file for each point. Instance segmentation masks, which represented the segmented seedlings, were included in each of these files. After generating all the masks, we merged them into a single mosaic raster file, as a visual representation of all the segmented seedlings from our single-point feature prompt.

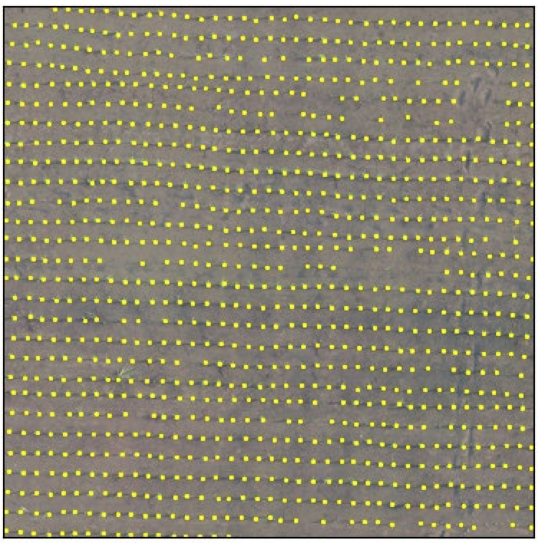

Figure 4. Rice seedlings annotated with points in yellow as input for SAM point prompt.

3.2.3 Automatic mask generation: As outlined by Kirillov et al. (2023), the automatic mask generator within SAM allows unsupervised segmentation by processing images without labeled data. SAM employs a transformer-based framework that analyzes entire images to generate segmentation masks. Its ability to automatically create high-quality masks without requiring annotated training data has significantly advanced computer vision workflows. The architecture includes an image encoder, which extracts hierarchical feature representations, and a mask decoder, which uses these features to produce accurate segmentation outputs (Kirillov et al., 2023; Wu & Osco, 2023). Despite other scenarios, in this scenario, there is no prompt encoder between image encoder and mask decoder (Wilk, 2023).

In all scenarios above, we employed a specialized tool called "SamGeo" to facilitate our geospatial analysis. This custom instrument is part of a larger module designed for comprehensive image segmentation. We opted to utilize the ViT-H SAM model, which represents the cutting-edge technology in the field, offering the most advanced capabilities available (Osco et al., 2023). Our approach involved utilizing the generate method of the SamGeo instance to perform general prompting. This process was straightforward, as it segmented the entire image and saved the results as an image mask file containing the segmentation masks. By personalizing this powerful toolset, we were able to simplify our analysis while benefiting from the sophisticated segmentation capabilities of the ViT-H SAM model.

3.2.4 Evaluation and comparison: Before conducting our accuracy assessment, we prepared the ground truth data by enhancing the greenness of the images using the ExGR index (N. Wang et al., 2022). As shown in Figure 5b, this index highlights seedlings and vegetation in green, while the background and other regions appear in shades of orange or red. This preprocessing step helped to improve the visibility of vegetation in the images. Following this enhancement, we applied a multi-OTSU thresholding technique to transform the color-enhanced images into binary masks (Figure 5c). The multi-OTSU method, which is an extension of the classic Otsu's thresholding algorithm, allows for the identification of multiple thresholds within an image.

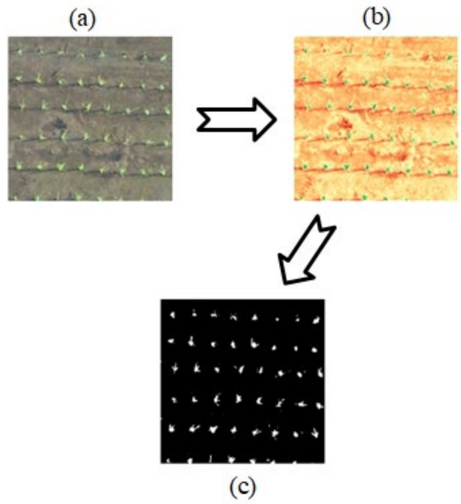

Figure 5. Ground truth mask generation process: (a) Original orthophoto; (b) ExGR index, differentiating rice seedlings with an emphasis on their greenness; (c) Binary mask as ground truth data.

This approach proved effective in distinguishing between different levels of vegetation density and other elements in the scene, ultimately leading to more precise binary masks for our subsequent analysis.

To evaluate the results, three criteria are considered. Processing time and visual results are important for comparing results. Moreover, for evaluating the accuracy of segmentation in our rice seedling detection task, we considered several key metrics. First, we treated the problem as a binary classification issue, where predictions could be categorized as either seedling or non-seedling. IoU (Intersection over Union) quantifies the overlap between predicted and ground-truth masks, providing insight into the model's precision in identifying seedlings. The Dice coefficient serves as a regional metric, accounting for potential class imbalances in the data. By focusing on these metrics, we aimed to obtain a comprehensive assessment of our model's performance in detecting and delineating rice seedlings accurately (Li et al., 2021; B. Wang et al., 2022). The evaluation index formula is as follows:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP+FP+FN}} \qquad (1)$$

$$\mathrm{Dice} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \qquad (2)$$

where TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

Finally, after computing the individual metrics, we took a comprehensive approach to assess our model's overall performance. We aggregated all the calculated metrics across the eight images in our dataset. Specifically, we calculated the mean Intersection over Union (mIoU) and mean Dice (mDice) scores for each scenario. This aggregation allowed us to obtain a balanced view of the model's performance across different images and scenarios. By considering these average metrics, we gained a clear indication of the segmentation quality and model effectiveness in detecting rice seedlings.

4. Results

In this section, we present and discuss the outcomes of our rice seedling detection experiment. We examine the visual results of the segmentation process and compare the effectiveness of different SAM prompts used in this task. By sharing both the visual outputs, analytical and performance comparisons, we aim to provide a comprehensive understanding of how SAM performs in this specific agricultural application.

4.1 Processing time

One criterion we focused on was processing time. We measured the total time required to initialize the SAM model and conduct the segmentation process for each model and prompt tested. All our experiments were performed in Google Colab, utilizing the T4 GPU for accelerated computation. This setup allowed us to efficiently compare various SAM configurations while maintaining consistent hardware specifications across all scenarios. By controlling for processing environment and hardware, we ensured that differences in performance could be attributed solely to the SAM implementation itself, rather than variations in computational resources. This controlled approach enabled us to gain valuable insights into the efficiency and effectiveness of different SAM prompts and configurations for the specific task of rice seedling detection. The results of the recorded processing time of 8 images are summarized in Table 1.

Scenario	Time (s)
Automatic mask generation	4020
Bounding box prompt	173
Point prompt	35

Table 1. Processing time recorded for each scenario

As it is clear from Table 1, the maximum processing time is related to automatic mask generation as the first scenario, which is about 25 times the bounding box and 120 times the point prompt. On the opposite, the point prompt method demonstrates remarkable speed, completing the segmentation process in under one minute. This stark contrast highlights the speed of the point prompt method, suggesting it as a preferable option for rapid image segmentation tasks. The bounding box prompt falls somewhere between these extremes, though still significantly faster than the automatic mask generation method. Overall, these processing times could help choose the appropriate SAM prompt for specific segmentation tasks, balancing accuracy requirements with computational efficiency.

4.2 Visual comparison

The results of the segmentation for different SAM models are shown in Figure 6.

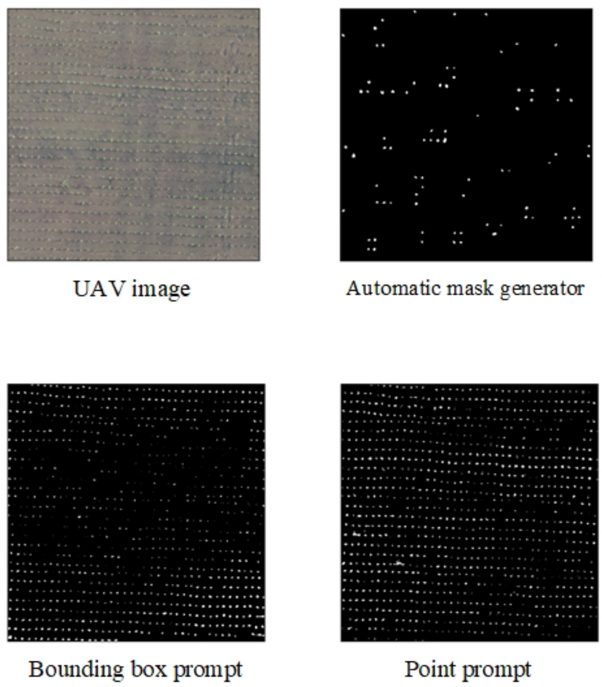

Figure 6. Results of the rice seedling segmentation using Automatic mask generation, bounding box and point prompt.

Upon examining the segmentation results, we observed notable differences in visual quality and accuracy across the three scenarios. The point prompt method yielded the most impressive results, demonstrating exceptional accuracy in

https://doi.org/10.5194/isprs-annals-X-G-2025-713-2025

identifying rice seedlings. Nearly all seedlings were correctly detected, with only a handful of small seedlings remaining undetected due to potential complications in extraction such as the size of the seedling. The bounding box approach showed high success in identifying most rice seedlings but failing to detect a higher percentage of seedlings than previous prompt. The automatic mask generation scenario fared poorly, with a significant portion of seedlings going undetected. Obviously, the poor visual performance will be reflected in the low values of the accuracy indexes, indicating a clear inefficiency in segmenting rice seedlings using this scenario.

Visual comparison with the original UAV images revealed that the point prompt method produced the most accurate segmentation maps, closely resembling the actual distribution of rice seedlings. The bounding box approach showed reasonable alignment with the ground truth, despite some discrepancies. Conversely, the automatic mask generation scenario failed to segment the seedlings, resulting in a poor visual match with the reference image.

4.3 Accuracy indexes

As mentioned earlier, we use the mIoU and mDice to evaluate the accuracy of the three considered scenarios. Figure 7 compares the accuracy indexes of these considered scenarios.

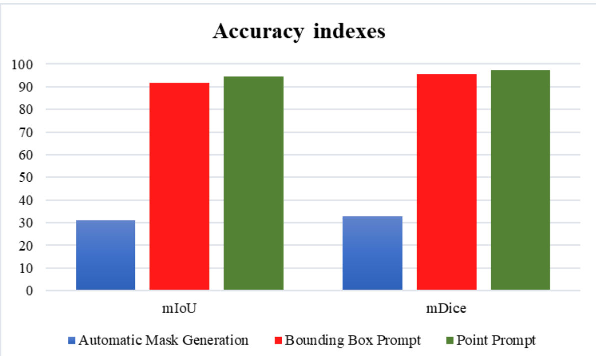

Figure 7. The evaluation and comparison of accuracy indexes across scenarios

Figure 7 reveals a clear ranking of segmentation accuracy across the three scenarios. The point prompt method emerged as the top performer, achieving a Dice score of 0.97. This exceptional accuracy demonstrates the model's ability to precisely identify rice seedlings. Following closely behind was the bounding box approach, which achieved a Dice score of 0.95. While slightly less accurate than the point prompt, this result still indicates strong performance in detecting rice seedlings. On the other hand, the automatic mask generation scenario lagged significantly behind, obtaining the lowest accuracy across all evaluation criteria. This substantial gap in performance underscores the inefficiency of this method for segmenting rice seedlings. These numerical results align perfectly with our visual analysis presented earlier. The superior performance of the point prompt method and the relative success of the bounding box approach were both anticipated based on our observations of the segmentation maps. The poor showing of the automatic mask generation scenario was also expected, given its visual shortcomings in accurately capturing the distribution of rice seedlings.

The results show the efficiency of SAM in segmenting rice seedlings. Also, these results underscore the importance of selecting the appropriate SAM prompt for specific agricultural applications, balancing between accuracy requirements and computational efficiency. The point prompt method emerges as the preferred choice for detecting rice seedlings, offering a good balance between detection accuracy and processing speed.

5. Conclusion

Smart farming relies heavily on the ability to accurately estimate crop yields. Given rice's status as the world's second most consumed food, developing efficient detection methods for rice seedlings is crucial. This study explored three scenarios for detecting rice seedlings using the SAM to investigate the potential of SAM-based techniques, for enhancing precision agriculture applications, especially in rice cultivation: automatic mask generation, bounding box prompt, and point prompt methods. Our analysis focused on three critical evaluation criteria: processing efficiency, visual assessment, and accuracy measurements. The results consistently demonstrated that the point prompt method outperformed the other scenarios across all three metrics. In terms of processing time, the point prompt method exhibited superior efficiency, completing the segmentation task significantly faster than the other approaches. Visually, the output maps generated by the point prompt method showed remarkable alignment with the actual distribution of rice seedlings in the UAV images, offering a clear advantage over the other scenarios. Furthermore, the point prompt method achieved the highest accuracy scores according to mIoU (94.57 %) and mDice (0.97) metrics. These outstanding performance indicators underscore the effectiveness of this approach in accurately identifying and segmenting rice seedlings.

References

Anuar, M.M.; Halin, A.A.; Perumal, T.; Kalantar, B. Aerial Imagery Paddy Seedlings Inspection Using Deep Learning. *Remote Sens.* **2022**, *14*, 274. https://doi.org/10.3390/rs14020274.

Hashim, N.; Ali, M. M.; Mahadi, M. R.; Abdullah, A. F.; Wayayok, A.; Kassim, M. S. M.; Jamaluddin, A. 2023. Smart farming for sustainable rice production: An insight into applications, challenges and future prospects. *Rice Science*, *31*(1). https://doi.org/10.1016/j.rsci.2023.08.004.

Islam, N.; Rashid, M.M.; Pasandideh, F.; Ray, B.; Moore, S.; Kadel, R. A Review of Applications and Communication Technologies for Internet of Things (IoT) and Unmanned Aerial Vehicle (UAV) Based Sustainable Smart Farming. *Sustainability* **2021**, *13*, 1821. https://doi.org/10.3390/su13041821

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; Dollár, P., 2023. Segment anything. arXiv preprint arXiv:2304.02643.

Li, M.; Wu, P.; Wang, B.; Park, H.; Hui, Y.; Yanlan, W. 2021. A Deep Learning Method of Water Body Extraction from High Resolution Remote Sensing Images with Multisensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 3120–3132. https://doi.org/10.1109/jstars.2021.3060769.

Osco, L. P.; Wu, Q.; Lemos, E. L.; Gonçalves, W. N.; Ramos, A. P. M.; Li, J.; Marcato, J.; The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot, *International Journal of Applied Earth Observation and Geoinformation*, Volume 124, 2023, 103540, ISSN 1569-8432, https://doi.org/10.1016/j.jag.2023.103540.

Tseng, H.-H.; Yang, M.-D.; Saminathan, R.; Hsu, Y.-C.; Yang, C.-Y.; Wu, D.-H. Rice Seedling Detection in UAV Images Using Transfer Learning and Machine Learning. *Remote Sens.* **2022**, *14*, 2837. https://doi.org/10.3390/rs14122837.

Wang, B.; Chen, Z.; Wu, L.; Yang, X.; Zhou, Y. 2022. SADA-Net: A Shape Feature Optimization and Multiscale Context Information-Based Water Body Extraction Method for High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 1744–1759. https://doi.org/10.1109/jstars.2022.3146275.

Wang, N.; Guo, Y.; Wei, X.; Zhou, M.; Wang, H.; Bai, Y. 2022. UAV-based remote sensing using visible and multispectral indices for the estimation of vegetation cover in an oasis of a desert. *Ecological Indicators*, *141*, 109155. https://doi.org/10.1016/j.ecolind.2022.109155

Wilk, J. 2023. *How to Use the Segment Anything Model (SAM) to Create Masks*. freeCodeCamp.org. https://www.freecodecamp.org/news/use-segment-anything-model-to-create-masks.

Wu, J.; Yang, G.; Yang, X.; Xu, B.; Han, L.; Zhu, Y. Automatic Counting of in situ Rice Seedlings from UAV Images Based on a Deep Fully Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 691. https://doi.org/10.3390/rs11060691.

Wu, Q; Osco, L. P, 2023. samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM). *Journal of Open Source Software*, 8(89), 5663, https://doi.org/10.21105/joss.05663.

Xiao, W.; Xu, S.; He, T. Mapping Paddy Rice with Sentinel-1/2 and Phenology-, Object-Based Algorithm—A Implementation in Hangjiahu Plain in China Using GEE Platform. *Remote Sens.* **2021**, *13*, 990. https://doi.org/10.3390/rs13050990.

Yang, M.-D.; Tseng, H.-H.; Hsu, Y.-C.; Yang, C.-Y.; Lai, M.-H.; Wu, D.-H. A UAV Open Dataset of Rice Paddies for Deep Learning Practice. *Remote Sens.* **2021**, *13*, 1358. https://doi.org/10.3390/rs13071358.

Appendix

The dataset used in this study was obtained from Rice Seedling Dataset, originally published in "A UAV Open Dataset of Rice Paddies for Deep Learning Practice" in Remote Sensing Journal of MDPI by Yang et al., 2021. The dataset is publicly available in the GitHub repository and can be accessed at the following link: https://github.com/aipal-nchu/RiceSeedlingDataset