Automatic Segmentation of Crops in UAV Images

A Project Paper

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Professional Studies

by

Yuanyuan Zheng

August 2023

ABSTRACT

Remote sensing imagery has been increasingly utilized in agricultural production due to its convenience and cost-effectiveness. However, traditional methods for crop segmentation require significant time and manual effort. Therefore, this research proposed the use of threshold segmentation and deep learning techniques to achieve automatic crop segmentation in UAV images and evaluated their performance. Specifically, this research utilized image threshold segmentation, a custom UNet network, Deeplabv3+ and segment anything model(SAM) with multiple prompts.

The results showed that the Intersection over Union (IoU) for threshold segmentation was 0.58. The IoU for UNet was 0.70, and for DeepLabV3+ it was 0.76. The IoU achieved by SAM with points prompt was 0.89, demonstrating superior crop segmentation performance. However, the masks generated using SAM automatic mask generation and a bounding box with a point prompt couldn’t segment crops effectively.

BIOGRAPHICAL SKETCH

Zheng Yuanyuan, born in the year 2000, is a passionate advocate for geographic spatial science, driven by the vision of making the world a better place through its application. Her journey in this field began in 2018 when she enrolled in the Geographic Information Science program at Anhui Normal University. During her academic years, she actively engaged in projects related to remote sensing, machine learning, and spatial analysis. For her undergraduate thesis, she explored the spatial correlations between PM2.5 and nighttime lights in China.

In 2022, Zheng Yuanyuan pursued a MPS degree in Integrated Plant Science at Cornell University. Specializing in geospatial applications, she delved deeper into the realms of machine learning, remote sensing, and GIS. Her research focus revolved around the automatic segmentation of crops in unmanned aerial vehicle (UAV) imagery, leveraging the potential of cutting-edge technology to enhance agricultural practices.

ACKNOWLEDGMENTS

First of all, I would like to express my heartfelt gratitude to my advisor, Prof. Yu Jiang, for providing invaluable academic guidance and support in both my scholarly pursuits and personal life. His counsel and assistance have been instrumental in shaping my academic journey.

I would also like to express my profound appreciation to my parents and friends, whose unwavering belief in me and unwavering encouragement have been a constant source of strength. Their emotional and material support have enabled me to steadfastly pursue my passion and forge my own path.

Lastly, I extend my sincere thanks to the entire faculty and staff of Cornell University. Their commitment to creating an enriching learning environment has allowed me to delve deeply into subjects that I am passionate about. Cornell has provided me with a wealth of knowledge and opportunities that have made a world of difference in my experience at Cornell.

To all those mentioned above, I am deeply grateful for your contributions to my academic and personal growth. Your support has been indispensable, and I am privileged to have you as part of my journey.

TABLE OF CONTENTS

1. Introduction

Agriculture is essential to the human society. As the world progresses, the integration of technology into agricultural practices has become more prevalent. Monitoring crop growth is important for optimizing management, pest and disease prevention, resource optimization. However, traditional methods such as human field assessment are labor-intensive, time-consuming, and resource-consuming. The use of unmanned aerial vehicles (UAVs) has proven to be a promising solution. By capturing remote sensing imagery, UAVs enable quick, convenient, and wide-scale monitoring of crop growth. Nevertheless, the traditional way of processing remote sensing images, i.e., visual interpretation, is not only demanding on operators but also slow in processing. Therefore, automatic crop segmentation of UAV images is of great significance to promote agricultural production.

Currently, there are several algorithms for automatic crop segmentation. Threshold segmentation, as a primitive method, is popular for its simplicity, low cost, and stable performance (Weszka, 1978). With the continuous development of internet technology, deep learning-based image segmentation methods have become increasingly advantageous. The key to the threshold segmentation algorithm is to find the optimal grayscale threshold based on certain criteria (Xiaonan Zheng et al., 2020).Meyer, Hindman, & Laksmi utilized ExG as an indicator for plant identification and proposed that the Red-Green-Blue (RGB) format may offer the most optimal segmentation criteria (Meyer, Hindman, & Laksmi, 1999). Onyango & Marchant combined R, G, and B channels with planting grids and probabilistic approaches to investigate the performance of segmentation algorithm with changing parameters (Onyango & Marchant, 2003). A.B. Payne employed RGB images and the YCbCr color space to extract features for eliminating or including pixels in a binary mask of the image. They further performed NDI

threshold segmentation, converted grayscale values, and Cr, Cb layers to successfully segment mangoes from the background (Payne et al., 2013). However,when the reflectance of different land cover classes does not vary significantly, it becomes challenging to effectively segment crops.

Computer vision includes object detection, image segmentation, object localization, and image classification. Among these tasks, image segmentation has become a prominent research focus (Huang, Zheng, & Liang, 2020). Semantic segmentation of remote sensing imagery pertains to pixel-level classification of land cover classes in high-resolution remote sensing images captured by satellites or unmanned aerial vehicles (UAVs) (Yuan, Shi, & Gu, 2021). UNet is a commonly used semantic segmentation model. Its architecture, characterized by encoder-decoder networks with skip connections, allows for precise delineation of object boundaries and efficient feature extraction (Ronneberger, Fischer, & Brox, 2015). Fawakherji et al. conducted a research on an agricultural robot that utilized UNet for vegetation segmentation in RGB images. Then, a deep Convolutional Neural Network (CNN) was used for crop/weed classification. This approach demonstrated good results across various environmental conditions (Fawakherji et al., 2019). Divyanth et al. used UNet in a two-stage approach for corn leaf disease recognition. The first stage involved segmenting corn leaves from the field, and the second stage is identifying the diseased regions on the corn leaves. UNet model performaned well in both stages(Divyanth, Ahmad, & Saraswat, 2023). Currently, more studies are focusing on improved versions of UNet. Kamath et al. proposed an ASPP-UNet network that enhanced row segmentation performance for maize crops. By integrating ASPP into the encoder and decoder, it can improve overall segmentation accuracy (Kamath et al., 2022).

The DeepLab series of semantic segmentation models proposed the Atrous Spatial Pyramid Pooling(ASPP) method, which enlarges the receptive field without changing the resolution, effectively fusing features from different levels and achieving excellent object boundary segmentation. DeepLabv3+ is an improvement over DeepLabv3 that enhances semantic segmentation results by adding a simple yet effective decoder module, which helps refine the segmentation results, providing more detailed and precise segmentation boundaries (Chen et al., 2018). DeepLabv3+ achieved 89% accuracy on the public dataset Pascal VOC 2012 without any post-processing. This outstanding performance has made DeepLabv3+ widely utilized for tasks such as crop segmentation, crop growth monitoring, and crop pest identification. Wu et al. explored four different backbones for DeepLabv3+ ,i.e. ResNet-50, ResNet-101, Xception-65, and Xception-71, and applied Medium Frequency Weight (MFW) and Uniform Weight (UW) assignation methods to mitigate data imbalance. The results showed that ResNet-101 combined with UW had the best segmentation performance, while ResNet-50 had the fastest segmentation speed (Wu et al., 2021). Moreover, some studies have made improvements to DeepLabv3+ based on their research objectives. Peng et al. enhanced the semantic segmentation of litchi branches by combining the DeepLabV3+ model with Xception depthwise separable convolutional features and reduced network parameters, which improved MIoU of 0.144 (Peng et al., 2020). Given the relatively small size of this research dataset, the ResNet-50 backbone is sufficient for crop segmentation.

Segment anything model is a semantic segmentation model released by Meta AI Research that does not require additional training. It was trained on a numerous natural image datasets. The model design and training demonstrate both efficiency and excellent performance in object segmentation. Ji et al. utilized SAM's Click, Box, and Everything modes for crop segmentation

in various agricultural scenarios. The results indicated that while SAM may not achieve optimal generalization in agricultural scenes, it can still achieve good segmentation results on images with relatively obvious plants (Ji et al., 2023). In this study, the combination of UAV imagery and plot polygons was employed, with each polygon containing only one crop, so a good segmentation result can be achieved.

In this research, the above models are investigated for automatic crop segmentation in UAV imagery to compare their performance.

2. Data

Two datasets were used in this study: 1) Data collection involved weekly flights using the DJI Matrice 600 drone over the Cornell Pathology vineyard. The data collection period is from mid-June to mid-August of 2022. The flights over the Cornell Pathology vineyard are part of a larger study that aims to develop multimodal, multi scale remote sensing tools for grape disease detection. The images were orthorectified and mosaicked in Agisoft Metashape. After the Metashape process, the bands were rearranged based on the wavelength order. band 1 and 2 are blue, band 3 and 4 are green, and band 5 and 6 are red, band 7 and band 8 are NIR1, band 9 and band 10 are NIR2. The coordinates of vineyard trellis posts was used to generate a geojson file containing polygon geometries for each panel of grapevines (1 panel= 3 or 4 vines). Then, the geojson was used to clip the orthomosaics for further processing and segmentation of the vine canopy. 2)During growing season in 2022, a high-cannabinoid hemp wide-spaced trial was managed to advance multimodal proximal sensing and high-throughput phenotyping applications for hemp breeding and production. In this pilot trial, the researcher focused on biomass yield and conducted weekly flights from June to August 2022 with a DJI M600 equipped with a RedEdge-MX Dual Camera System. The resultant multispectral images were orthorectified and mosaicked

with Agisoft Metashape. The processed images were then used to create bounding boxes and polygon shape files.

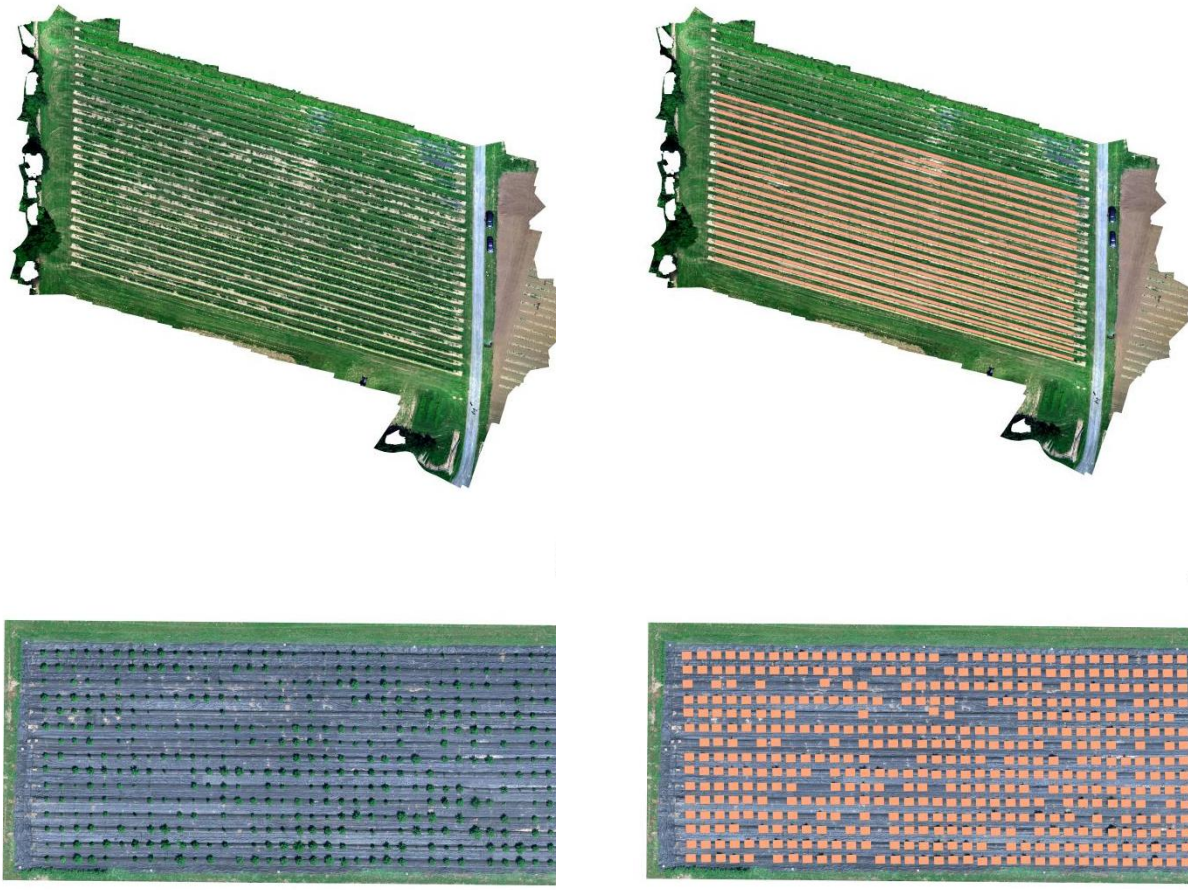


Figure 2-1 UAV images and polygons

3. Method

Before performing automatic segmentation, preprocessing of the drone imagery is necessary. Firstly, the imagery needs to be reprojected to the UTM coordinate system, and images and polygons were rotated using the same center. Subsequently, the images were cropped using the corresponding polygons. The drone imagery was then scaled to the same size before inputting UNet and DeepLabv3+. ArcGIS was utilized to create the training dataset for UNet and

DeepLabv3+ models. While, for the Segment Anything model, crop segmentation was conducted using synthetic RGB imagery generated from the drone imagery.

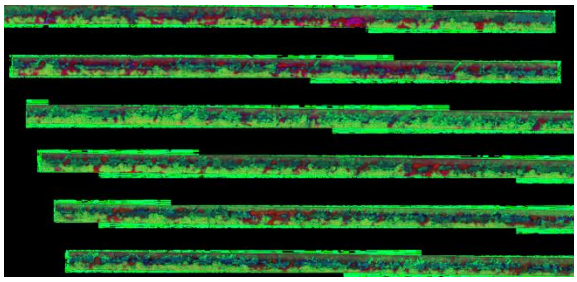
3.1. Threshold segmentation

Threshold segmentation is an algorithm that sets a threshold on the pixel values of an image to divide it into target object and background (Zhu et al., 2007).

In this study, three threshold segmentation algorithms were used:

1. NDVI Thresholding: The calculation formula of the Normalized Difference Vegetation Index (NDVI) is (NIR - R)/(NIR + R), which ranges between -1 to 1. NDVI utilizes the characteristics of NIR and Red bands of the vegetation spectrum and can effectively enhance vegetation information, which makes it a suitable index for crop segmentation. Pixels with NDVI greater than 0.5 were considered as crops.

2. YCbCr thresholds: I converted the UAV images to RGB, HSV, and YCbCr color spaces. The results showed that the distinction between crops and soil is higher in the Cr and Cb layers. In the YCbCr color spaces, Y represents the luma component, CB represents the blue-difference component, and CR represents the red-difference component. In this study, pixels with Cb < 108 and Cr < 117 were identified as crops.



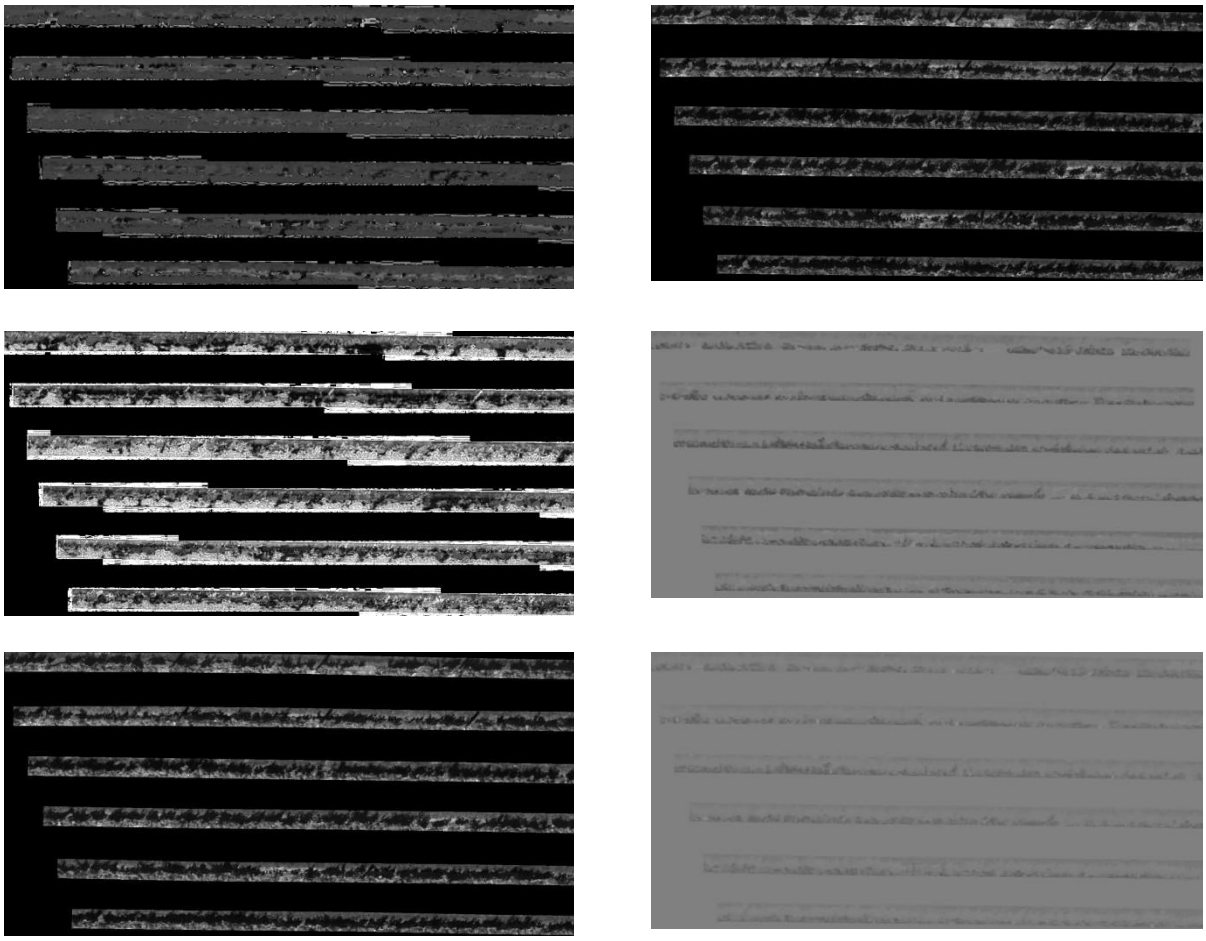


Figure 3-1 HSV image(left) and YCbCr image(right)

Pixels that satisfy all three threshold segmentation conditions are recognized as crops. Then, an 8*8 kernel was used to eliminated hole in the masks.

In this study, the threshold image G(x,y) can be define:

$$G(x,y)\begin{cases}0, NDVI \leq 0.5 \; and \; Cb \geq 108 \; and \; Cr \geq 117 \\ 1, NDVI > 0.5 \; and \; Cb < 108 \; and \; Cr < 117\end{cases}$$

3.2. UNet

UNet is a widely used semantic segmentation model, consisting of an encoder and decoder that form a U-shaped architecture. At each step of the contracting path, UNet reduces the spatial dimensions by half while doubling the number of feature channels.Conversely, in the expansive

path, the number of feature channels is halved, and the spatial dimensions are doubled. Additionally, UNet incorporates skip connections, allowing for smoother gradient flow during backpropagation, and more effectively in performing semantic segmentation (Ronneberger, Fischer, & Brox, 2015). The UNet network constructed in this study has four encoders and decoders and supports multi-band image inputs.

3.3. DeepLabV3+

DeepLabV3+ is a typical network for semantic segmentation. It adopts a residual network as the underlying backbone network and incorporates the Atrous Spatial Pyramid Pooling (ASPP) module, which allows for encoding multi-scale contextual information to avoid information loss. An encoder-decoder architecture is added to restore spatial information and optimize boundary segmentation. Furthermore, DeepLabV3+ introduces Modified Aligned Xception and Atrous Separable Convolution, resulting in a faster and more powerful network that enhances segmentation accuracy (Chen et al., 2018).

3.4. Segment Anything Model

The Segment Anything Model (SAM) utilizes a powerful image encoder to compute image embeddings, along with a prompt encoder for embedding prompts and then predicts the segmentation mask by combining the two sources of information in a lightweight mask decoder (Kirillov et al., 2023). SAM demonstrates excellent performance under the zero-shot learning regime and supports various types of prompts.

In this study, SAM's performance was evaluated using different types of prompts, including a bounding box and a point, a set of points, and automatically generated masks:

1. I used polygon boundaries as bounding boxes and the center points of the polygons as foreground points to predict masks.

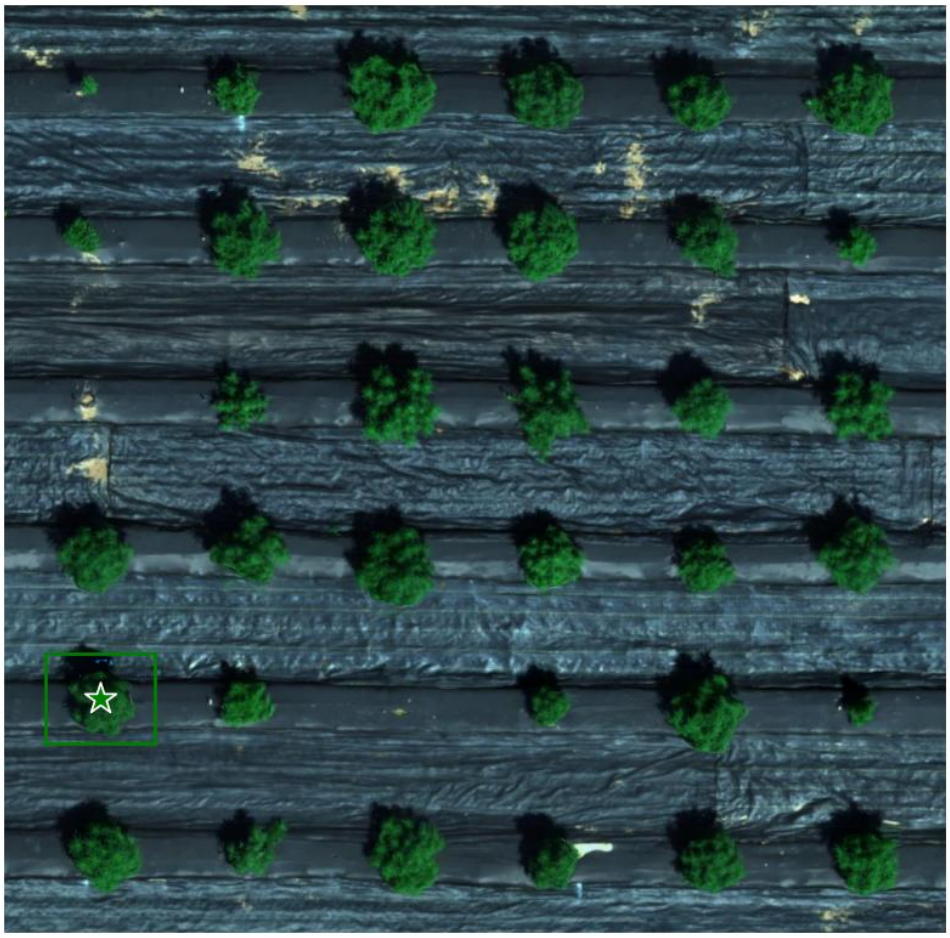


Figure 3-2 Bounding box and center point prompt

2. SAM supports automatic mask generation, segmenting all possible objects in the image, and generating multiple masks. Among the generated masks, the final output was selected based on the following conditions: 1) the mask has a relatively large area; 2) the mask area is <9000, to exclude masks that cover almost the entire image, which SAM tends to generate frequently; 3) the average Green value and NDVI of the segmentation image are calculated, and masks with Green>90 and NDVI>0.7 are chosen.

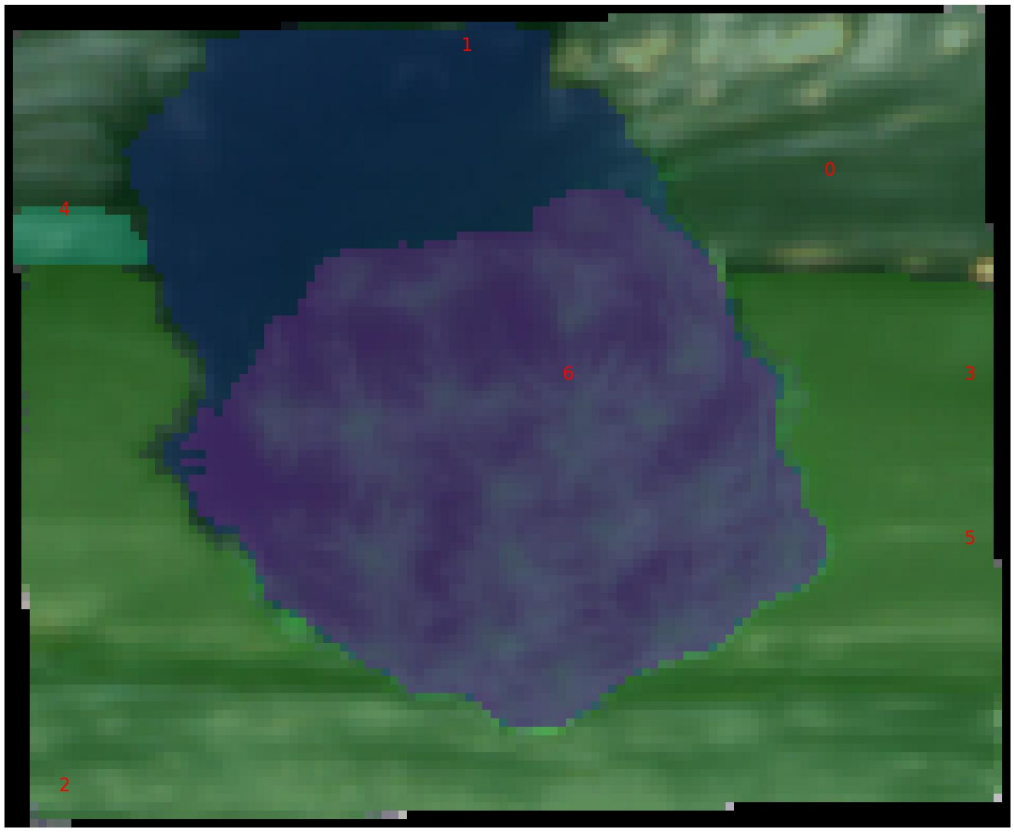


Figure 3-2 Automatic mask generation

3. Using the image center point as the foreground point and three points near the image boundaries as background points to generate masks. The masks must meet the following criteria: 1) the mask has a high score; 2) the mask area is <9000; 3) since the generated masks may contain shadows, the average Green value and NDVI of the segmentation image are calculated. Masks with Green>90 and NDVI>0.7 are chosen to exclude masks that contain shadows.

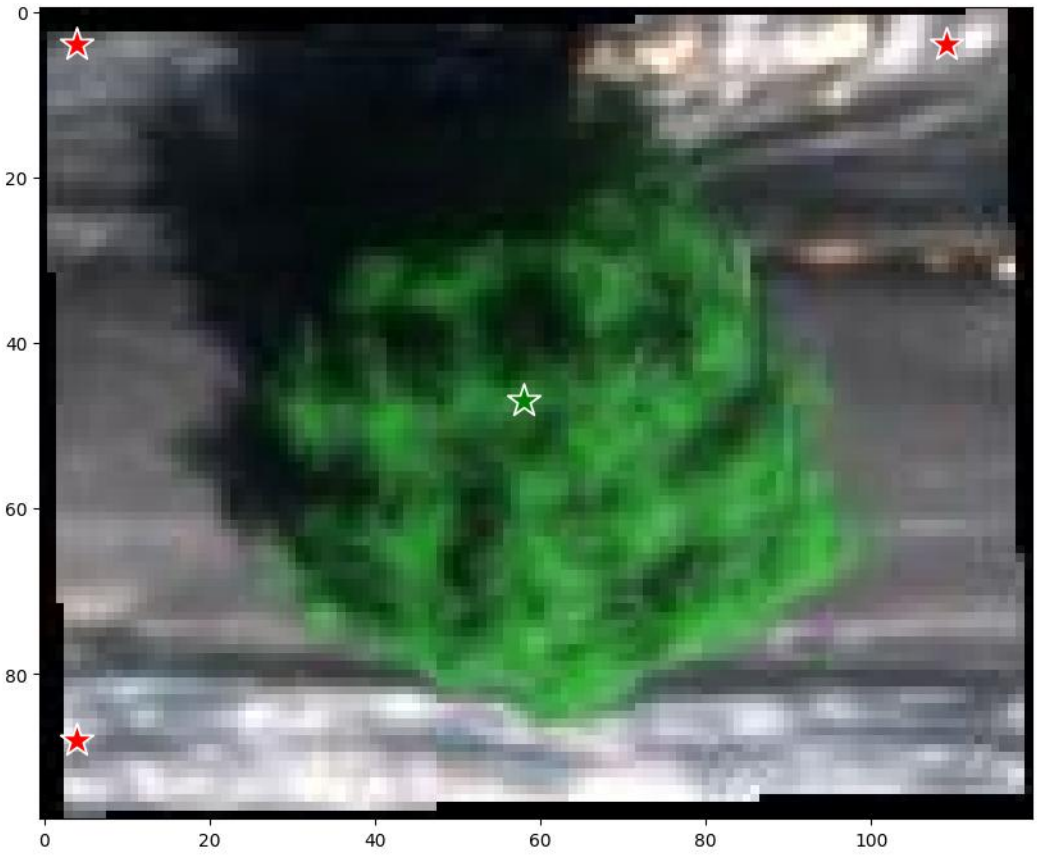
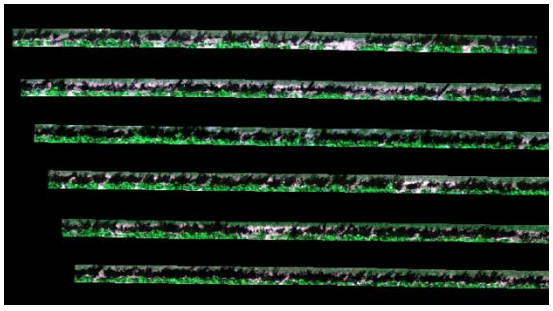
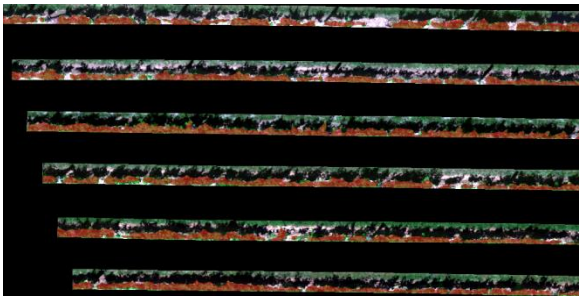


Figure 3-3 Points prompt

4. Results





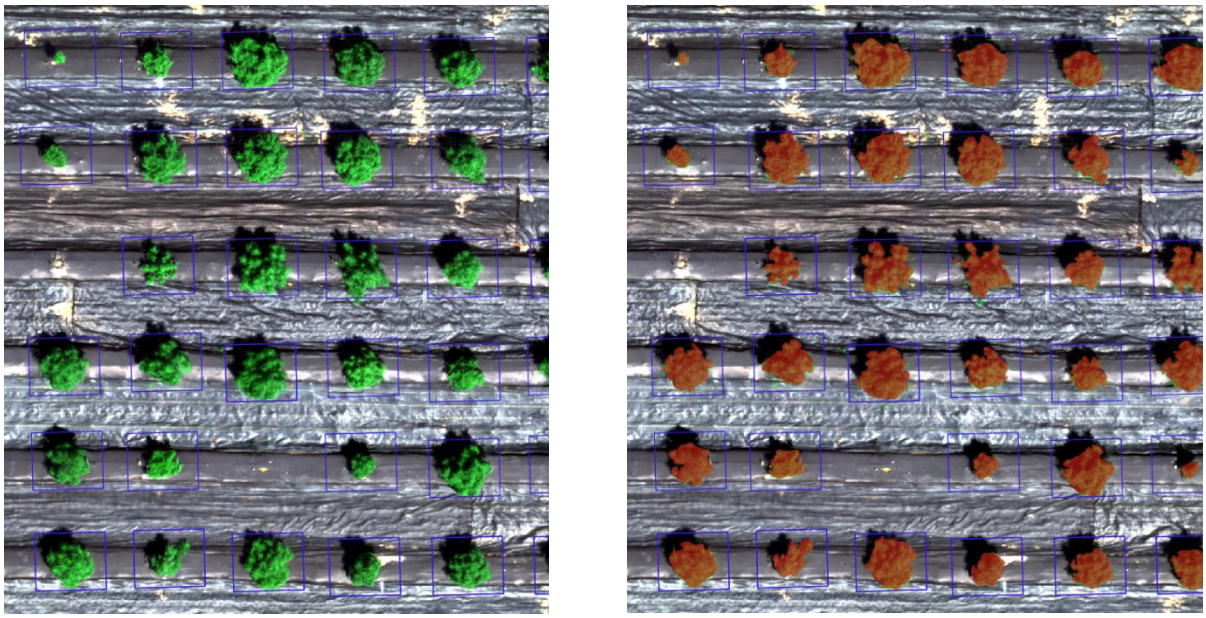
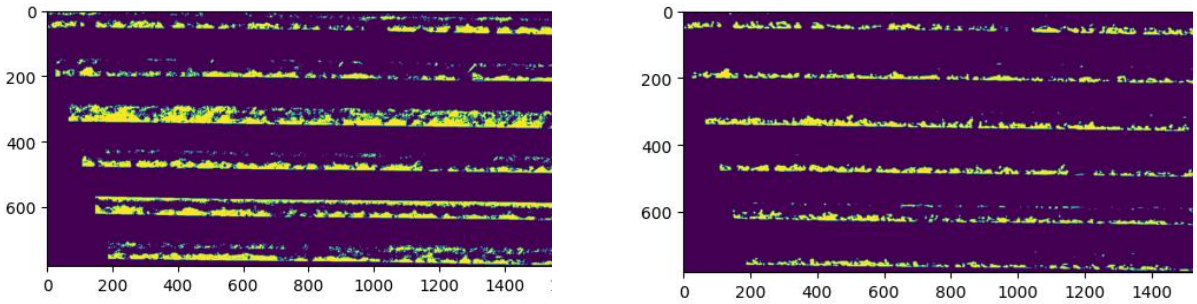


Figure 4-1 Test dataset and ground truth

UNet and DeepLabV3+ models were evaluated on Figure 4-1 test dataset to calculate the Intersection over Union (IoU). IoU is a widely used metric in computer vision, which is calculated as Area of Overlap / Area of Union between the predicted segmentation and the ground truth.

As the Segment Anything model lacked sufficient training on agricultural related images, it was utilized on a simpler dataset. This datsset contains one individual crop plant within each polygon. SAM also used IoU to evaluate its performance.

4.1. Threshold segmentation



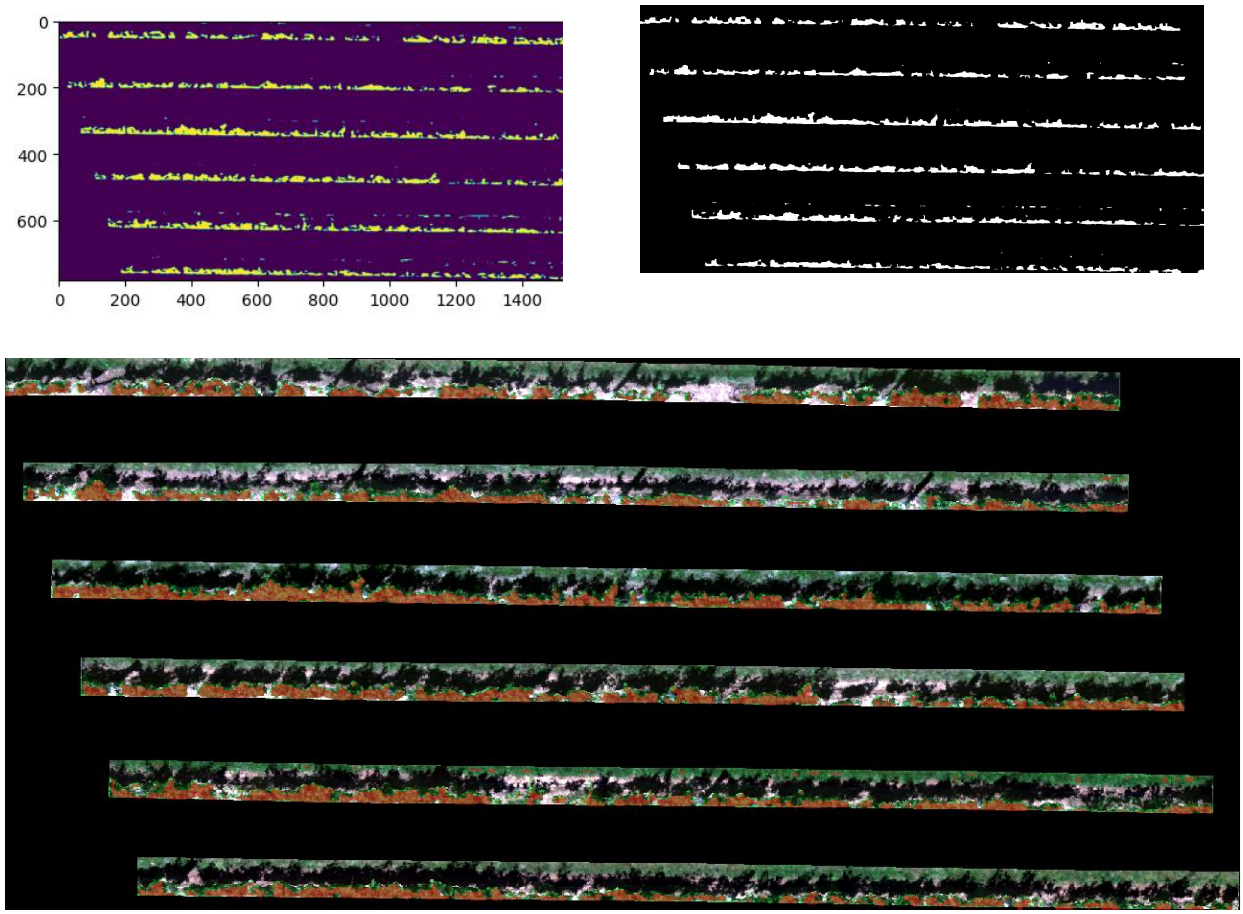


Figure 4-2 Threshold segmentation result (NDVI threshold segmentation, Cb threshold segmentation, Cr threshold segmentation, holes elimination result and final threshold segmentation result)

The IoU for threshold segmentation is 0.58. The result showed that it contained a small amount of weeds. This is because the NDVI,Cb layer and Cr layer of some weeds are very similar to the crop, so the threshold segmentation algorithm could not segment them effectively. Additionally, the threshold segmentation algorithm did not correctly identify crop pixels that were obscured by shadows. Besides, it requires manual selection of thresholds. Thus, more intelligent deep learning based crop segmentation methods need to be introduced.

4.2. UNet

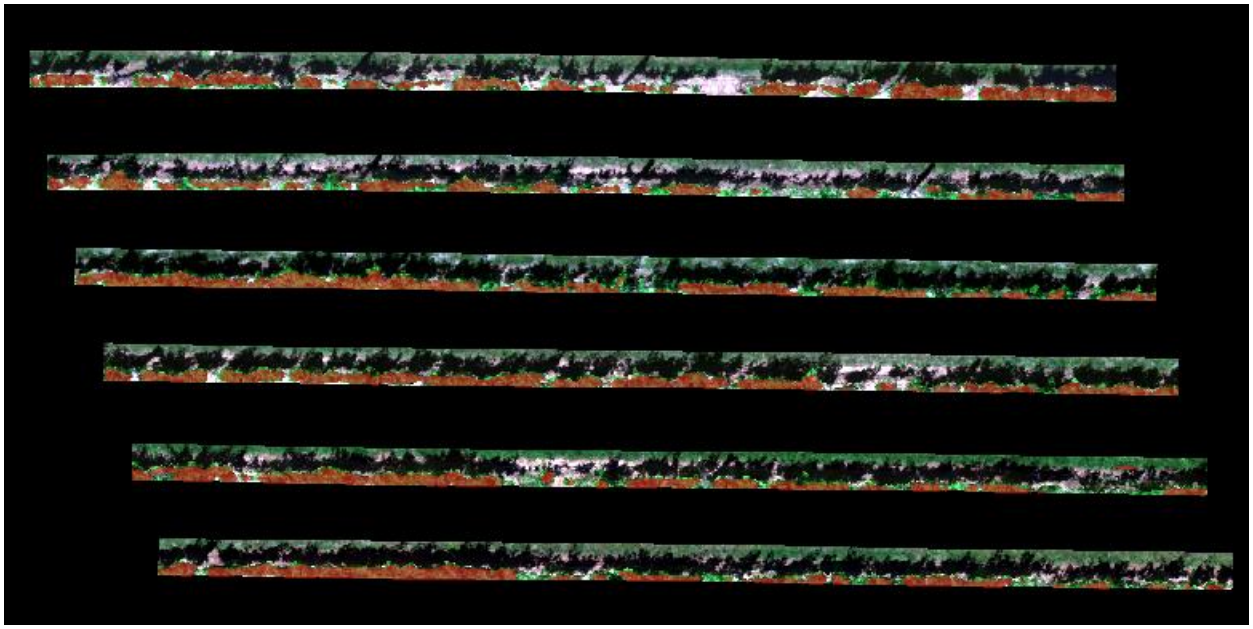


Figure 4-3 UNet segmentation result

In this study, a UNet network with four encoders and decoders was constructed, supporting multi-band image inputs. The IoU for UNet is 0.70. It can be seen that the segmentation results of UNet are band-like with smooth edges, which may be attributed to the characteristics of the training dataset. What's more, UNet demonstrates the capability to segment crops obscured by shadows. However, it struggles to effectively segment crop with less distinct features.

4.3. DeepLabV3+

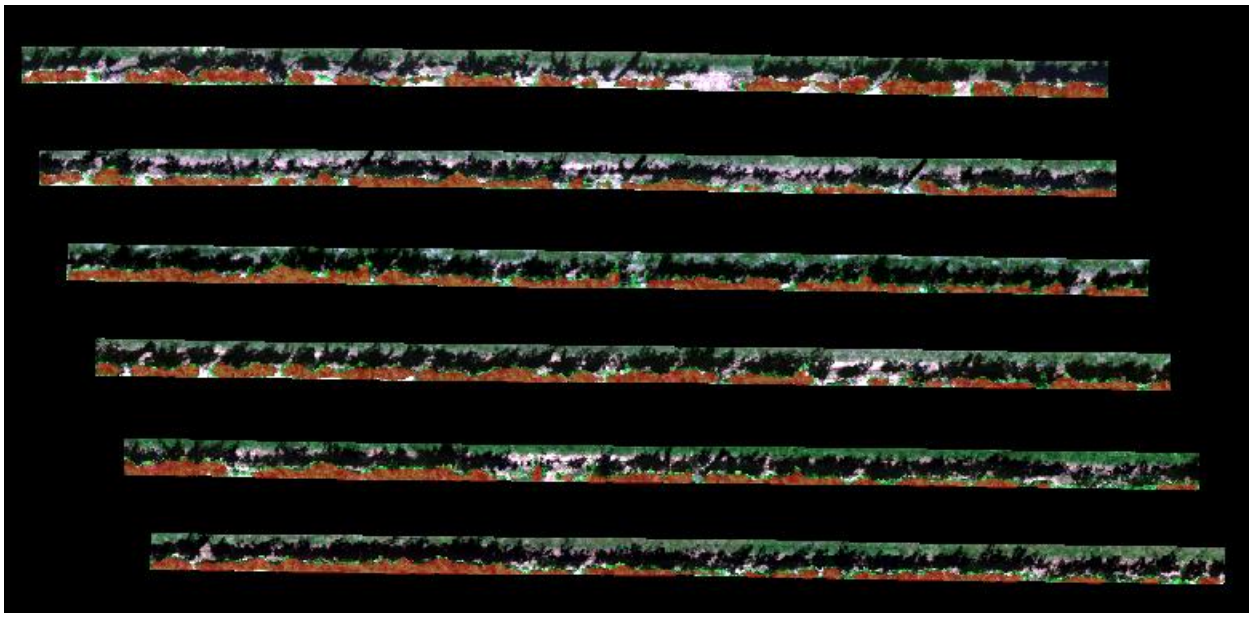


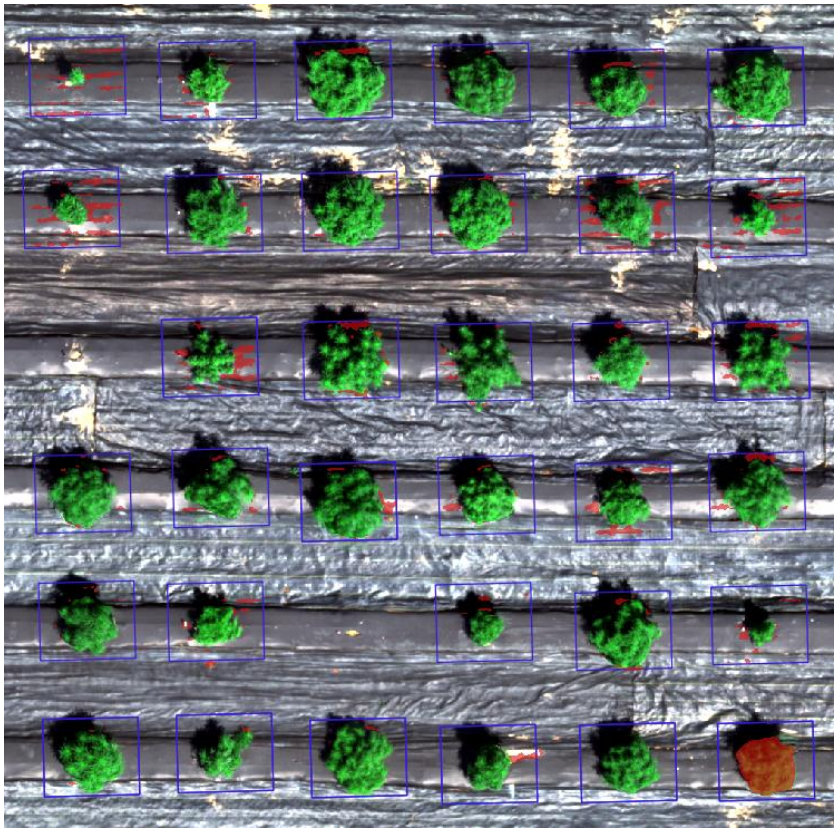
Figure 4-4 DeepLabV3+ segmentation result

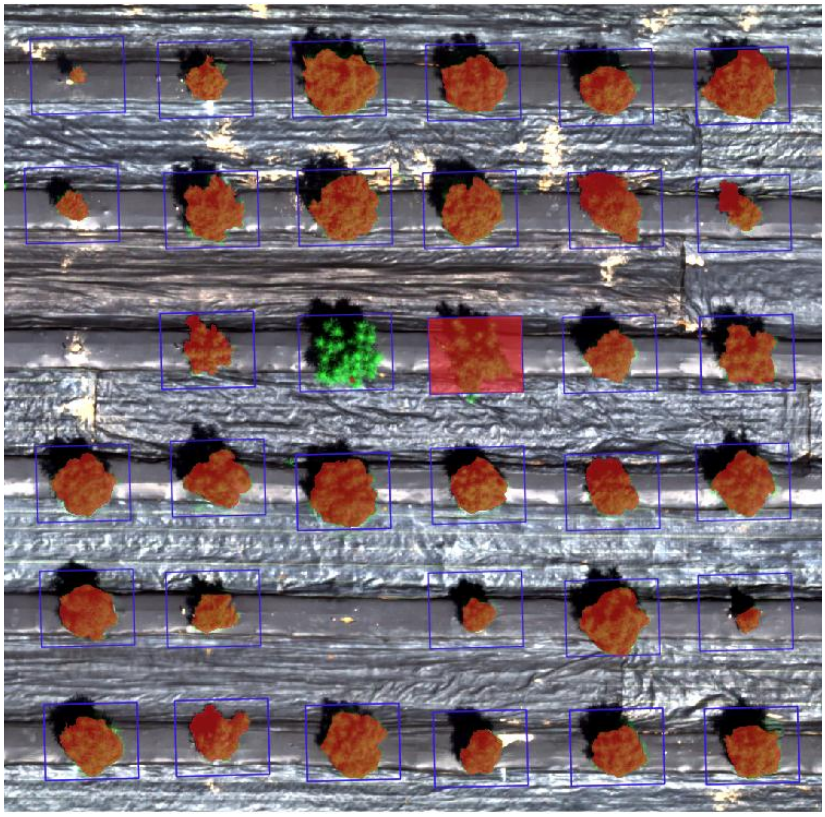
In this study, decoder of the DeepLabV3+ model was ResNet50 architecture and encoder weights initialized using pre-trained weights from the ImageNet dataset. The activation function used for the final segmentation output was the sigmoid.

The IoU for DeepLabV3+ is 0.76, an improvement over UNet. The segmentation results of DeepLabV3+ are similar to those of UNet, but with the advantage of capturing more details in the segmented crop.

The segmentation results demonstrates the ability of DeepLabV3+ to capture finer details, thus improving the accuracy in identifying and delineating crops.

4.4. Segment anything model





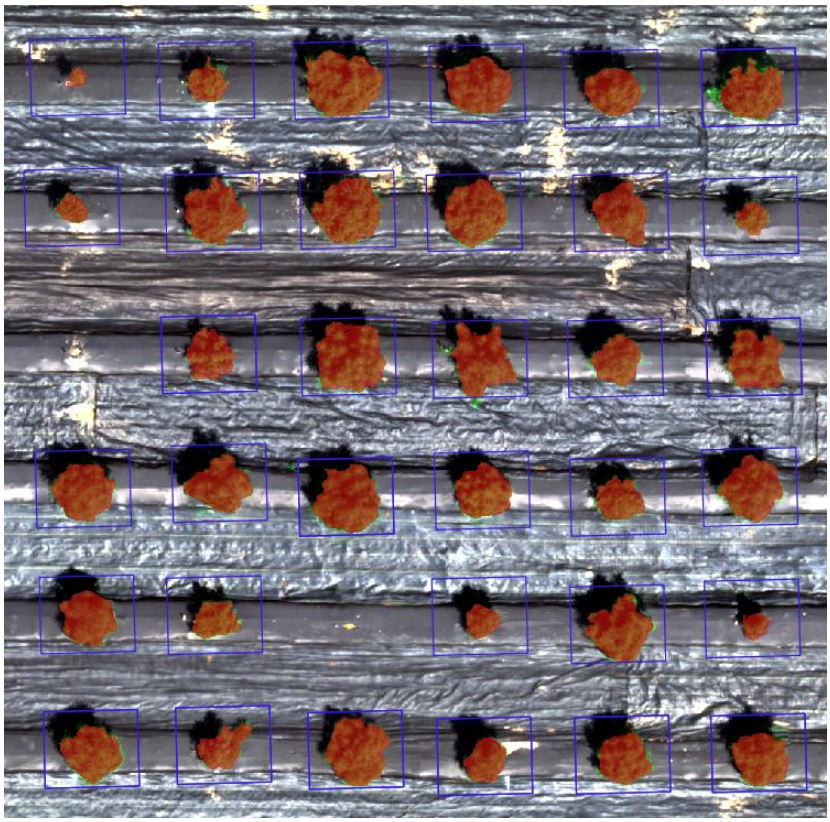


Figure 4-5 SAM with the bounding box and point prompt segmentation result, SAM with automatic mask generation result and SAM with points prompt segmentation result

It can be seen that the masks generated using bounding boxes and center points are very poor and problematic for successive data analysis. The performance of SAM automatically generated masks has improved significantly compared to bounding boxes, but plants with complex canopy morphology still presented challenges for the SAM model to achieve a satisfactory performance. The results generated using point prompts were excellent, segmenting the crop almost completely and also segmenting the complex vegetation edges well with an IoU of 0.89.

5. Discussion

SAM with points prompt had the best crop segmentation performance with IoU of 0.89, followed by DeepLabV3+ with IoU of 0.76, then UNet with IoU of 0.70, and finally threshold segmentation with IoU of 0.58. In terms of computation speed, threshold segmentation was the fastest, followed by SAM, and lastly UNet and DeepLabV3+. It demonstrated that while SAM may not meet the crop segmentation requirements in all scenarios, it performs well in segmenting relatively simple images and does not require manually labeled. Overall, the application prospects of SAM are promising.

There is still room for improvement in this study. Firstly, there is a limitation in the training dataset for both the UNet and DeeplabV3+ models, which is not extensive enough to meet the requirements of crop segmentation across all time periods. Secondly, in the case of SAM, crops cannot be segmented occasionally, and thresholds still need to be manually specified when selecting masks.

6. Conclusion

In this study, I compared the performance of different crop segmentation algorithms and the results showed that:

1. Threshold segmentation is the fastest method but has the lowest segmentation accuracy.

2. DeepLab V3+ and UNet segmentation results were better but they had the highest operational complexity due to the need for manual labeling. However, they can be applied to more scenarios with custom training dataset.

3. SAM has the highest accuracy with a fast speed and can be used without additional training. Nevertheless, it is only suitable for segmenting plants with simple canopy morphology.

Besides, different results are generated using different prompts, and only SAM with points prompt got a good segmentation performance in this study.

The above algorithms have their own advantages and disadvantages, and it is important to choose the appropriate algorithm based on the actual needs.

References

1. Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." Proceedings of the European conference on computer vision (ECCV). 2018.

2. Divyanth, L. G., Aanis Ahmad, and Dharmendra Saraswat. "A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery." Smart Agricultural Technology 3 (2023): 100108.

3. Fawakherji, Mulham, et al. "Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation." 2019 Third IEEE International Conference on Robotic Computing (IRC). IEEE, 2019.

4. G.E. Meyer, T. W. Hindman and K. Laksmi, "Machine vision detection parameters for plant species identification", Proc. SPIE, vol. 3543, pp. 327-335, Jan. 1999.

5. Huang P., Zheng Q., Liang C. A survey of image segmentation methods. Journal of Wuhan University (Natural Science Edition), 2020, 66(06): 519-531.

6. Ji, Wei, et al. "Segment anything is not always perfect: An investigation of sam on different real-world applications." arXiv preprint arXiv:2304.05750 (2023).

7. Kamath, Radhika, et al. "Classification of paddy crop and weeds using semantic segmentation." Cogent engineering 9.1 (2022): 2018791.

8. Kirillov, Alexander, et al. "Segment anything." arXiv preprint arXiv:2304.02643 (2023).

9. Onyango, Christine M., and J. A. Marchant. "Segmentation of row crop plants from weeds using colour and morphology." Computers and electronics in agriculture 39.3 (2003): 141-155.

10. Payne, Alison B., et al. "Estimation of mango crop yield using image analysis–segmentation method." Computers and electronics in agriculture 91 (2013): 57-64.

11. Peng, Hongxing, et al. "Semantic segmentation of litchi branches using DeepLabV3+ model." IEEE Access 8 (2020): 164546-164555.

12. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

13. Weszka J S. A survey of threshold selection techniques[J]. Computer Graphics & Image Processing, 1978,7(2):259-265.

14. Wu, Zhenchao, et al. "Segmentation of abnormal leaves of hydroponic lettuce based on DeepLabV3+ for robotic sorting." Computers and Electronics in Agriculture 190 (2021): 106443.

15. Yuan X., Shi J., Gu L. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Systems with Applications, 2021, 169: 114417.

16. Zheng, X., Yang, F., and Li, F. A Review of Crop Image Segmentation Algorithms." Collection 19. 2020.

17. Zhu, Shiping, et al. "An image segmentation algorithm in image processing based on threshold segmentation." 2007 third international IEEE conference on signal-image technologies and internet-based system. IEEE, 2007.