

Segmentation of *Arabidopsis thaliana* Using Segment-Anything

Landon G. Swartz

Dept. of Electrical Engineering and Computer Science
University of Missouri
Columbia, MO, USA
lgsm2n@umsystem.edu

Suxing Liu

School of Plant Science
Arizona University
Tucson, AZ, USA
suxingliu@arizona.edu

David Mendoza Cozatl

Dept. of Plant Science and Technology
University of Missouri
Columbia, MO, Missouri
mendozad@missouri.edu

Kannappan Palaniappan

Dept. of Electrical Engineering and Computer Science
University of Missouri
Columbia, MO, USA
pal@missouri.edu

Abstract—Speedy Measurement of *Arabidopsis* Traits (SMART) is a high-throughput phenotyping pipeline that processes images to observe leaf-specific traits. The current pipeline utilizes a k-means clustering segmentation method. This method has limitations in some cases where leaves display dynamic change in color and shape. Segment-Anything, a new foundational model in segmentation, offers a new method of high-resolution segmentation of complex geometries present in plant systems. Here we present a new method for segmentation of plant using Segment-Anything and Grounded-Dino. A method of obtaining individual leaves was implemented to describe the advantages and limitations of the prompt-based method of deep learning models. An analysis of the segmentation results from Segment-Anything demonstrates that this is a powerful method for providing statistically valuable data for biological insights into novel plant traits under nutrient stress.¹

Index Terms—Segmentation, Segment-Anything, leaves, SMART, trait analysis, high-throughput phenotyping

I. INTRODUCTION

Nutrient stress within plants displays itself with dynamic changes in structure and colors throughout the entire plant and in specific leaves at different ages [1]. Speedy Measurement of *Arabidopsis* Traits (SMART) is a high-throughput phenotyping processing pipeline that processes images of *Arabidopsis thaliana* to observe leaf-specific traits, including area, color, and shape, for researchers to understand genome-phenome relationships during nutrient limitation. The pipeline currently utilizes a static k-means clustering segmentation algorithm. Due to the dynamic nature of nutrient stress, the segmentation often struggles to provide consistent and reproducible segmentation results that impact downstream trait analysis tasks, as seen in Figure 1.

Segment-Anything Model, or SAM, is a recent foundational model from Meta that offers state-of-the-art results in zero-shot segmentation [2]. The model is designed to be generalized

¹U.S. Government work not protected by U.S. copyright. This work was supported by the National Science Foundation, Grant Numbers: DEB 2217322, MCB 2224839, 1845760

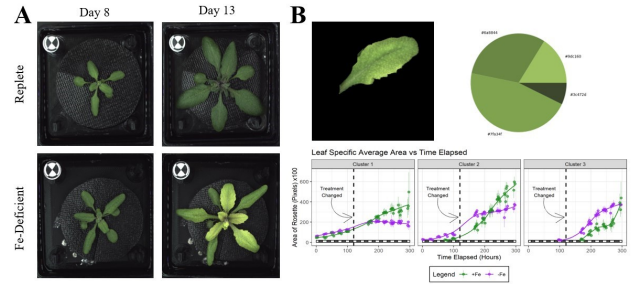


Fig. 1: (A) Plants under nutrient stress display dynamic changes in both shape and color in different aged leaves at different points. Here we show the changes between two plants under iron deficiency. (B) Trait analysis in the SMART pipeline can capture both shape and color but depend on segmentation for trait measurement. Trait measurement in precision agriculture requires multiple samples for statistical significance and can be limited by poor trait measurement.

for any research field or image. In turn, SAM can emerge as a method for tracking the dynamic changes of plants under nutrient stress to improve plant phenotyping pipelines.

In this work, we explore the use of SAM and its extensions in the field of plant phenotyping. We show the powerful ability of SAM to identify plants from the background with prompts as simple as a bounding box and text. However, SAM is error-prone and requires refinements that can be achieved through classical image processing techniques. We also explore the abilities of SAM to segment the parts of an object, such as each leaf at a specific time point. SAM has the potential of accelerating trait analysis in underserved research areas, such as plant phenotyping.

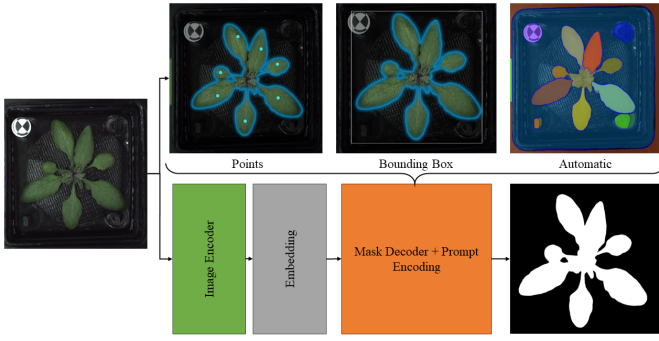


Fig. 2: Segment-Anything has three components. The image is first encoded using an MAE-ViT encoder. A prompt encoder uses convolutions summed element-wise to be embedded with the image. A mask decoder maps the image and prompt embedding to a mask output.

II. RELATED WORK

A. Plant Segmentation

Extracting accurate and quantitative measurements of plants requires robust segmentation of the plant as a whole and individual organs, like leaves [3]. However, plant segmentation is difficult due to changes in shape appearance and characteristics while growing, occlusions, and variability of leaves [4]. Most methods developed for plant segmentation started by utilizing classical segmentation methods, especially within the last major challenge for leaf segmentation and counting at CVPR 2014 [4]. More recent methods have integrated deep learning techniques, like recurrent networks or attention [5], and convolutional neural networks [6], [7]. Our work focuses on utilizing and fine-tuning general-purpose foundational models in vision for plants to improve the results of trait measurement.

B. Segment-Anything

Segment-Anything (SAM) is a foundational model in segmentation from Meta. It is designed to be a prompt-based segmentation system with zero-shot generalization to unknown objects and images, without further training [2]. The model takes inspiration from NLP by using next token prediction for model pre-training and to solve diverse downstream tasks via prompt engineering [8]. The model contains three components: an image encoder, prompt encoder, and mask decoder (Fig 2). The image encoder is a Masked AutoEncoder (MAE) pre-trained Vision Transformer (ViT) that processes high-resolution image inputs. The prompt encoder takes in sparse, including points, bounding boxes, text, and dense, such as masks, inputs to be embedded using convolutions and summed element-wise with the image embedding. The mask decoder maps the image embedding, prompt embedding, and an output token to a mask. The model is trained on a 1.1B mask data set called SA-1B, one of the first massive scale segmentation data sets [2]. Many extensions have been developed for SAM for integration with other models. Grounding-DINO extends SAM

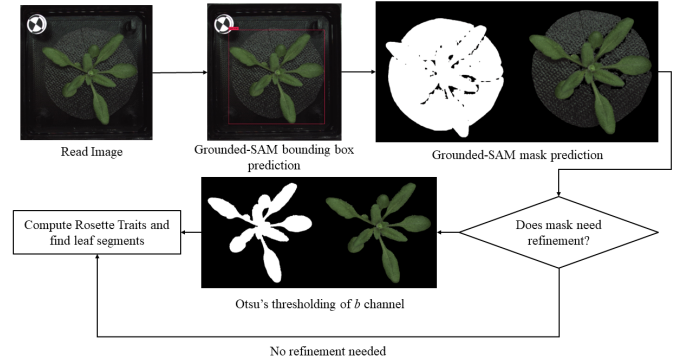


Fig. 3: To segment the entire plant from the background, we utilize Grounding-SAM to provide a bounding box prediction for SAM. If the mask prediction requires refinement, Otsu's threshold of the b channel allows for quick refinement to remove background noise that is not apart of the plant.

to take text prompts by providing bounding box predictions from text input for input into SAM [9].

III. METHODS

While the zero-shot capabilities for SAM are powerful, some tasks require unique prompt engineering to elicit the desired responses. Segmenting the entire plant from the background utilizing Grounded-SAM and refining the prediction when necessary. After segmenting the whole plant, each leaf is individually segmented using a combination of classical methods to provide point prompts for SAM.

A. Rosette Workflow

Segmentation of the entire plant is done with Grounding-SAM. A prompt of “green plant” is passed to Grounding-SAM to find the plant in each image. However, some mask predictions require refinement. One of the major drawbacks of SAM is that bounding-box prediction can include the background of an object along with the object. To determine if the mask needs refinement, we take the sum of the pixel colors. If the sum of the pixel colors is below a threshold, then the black background requires refinement. Otsu's threshold of the b channel of the $L*a*b*$ space finds the plant from the background of the initial mask.

B. Leaf Workflow

Segmenting the parts of an object using SAM requires unique prompt engineering. Segmenting the individual leaves from the entire plant starts with finding the canny edges of the a channel of the $L*a*b*$ space. This allows the edges between each leaf to be found. The edges are dilated and inverted to create individual sections of the leaves. The center points of active contours are found and passed to SAM for mask prediction. Some duplicates are produced by this method and are filtered out.

C. Evaluation Metrics

We utilize two metrics to determine the accuracy of the segmentations. The first is the Jaccard Index of the predicted mask and ground-truth mask [10]. As shown in equation 1, A is the predicted mask of an image while B is the ground-truth mask.

$$J(A, B) = \frac{|A \cup B|}{|A| + |B|} \quad (1)$$

Another metric we will use to evaluate the segmentations is standard error of trait analysis. One of the most important downstream consequences of improving the segmentations is that the trait analysis will be improved for statistical inference by researchers. The standard error is calculated as the standard deviation, σ , over the number of independent n samples:

$$\sigma_x^- = \frac{\sigma}{\sqrt{n}} \quad (2)$$

A lower standard error represents a more accurate and statistical significant measurement of a trait.

IV. EXPERIMENTAL RESULTS

We first observe the effects of our method with SAM on the extraction of the foreground plant from the background in two data sets. The first data set is an iron-deficiency experiment where four samples were under replete conditions and four samples were under iron-deficiency conditions using the OPEN leaf system. The second data set is the CVPR Leaf Counting Challenge to observe how the method translates to other experiment parameters in images. We found that our method, and other methods using SAM, are application-specific due to the prompt engineering involved. Lastly, we observe our ability to observe leaf-specific traits in using the first data set again.

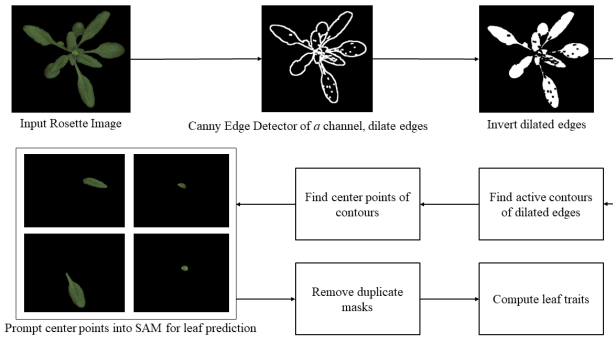


Fig. 4: Finding the parts of an object requires unique prompt engineering. Using a combination of canny edge detectors, thresholding, and active contours allow us to find the center points of leaves to pass to SAM.

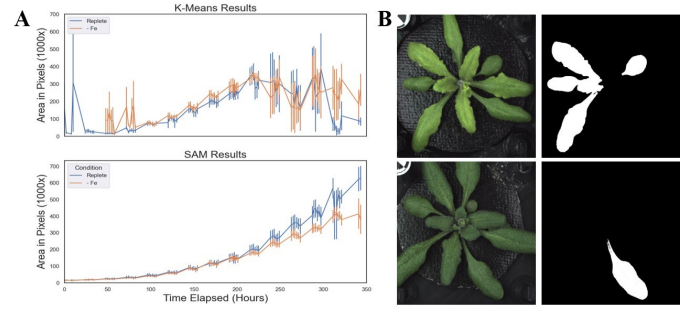


Fig. 5: (A) Trait analysis using our Grounded-SAM algorithm produces more stable standard error throughout the entire time series, especially at the end of experiments. (B) Examples of where the k-means algorithm struggles in the later stages of plant growth.

A. Rosette Segmentation

The task most simplified utilizing SAM was extracting the plant from the background of the image. However, the method for refinement in our work was found to not translate well between data sets. Therefore, the prompt engineering and prediction refinement for SAM are application-specific.

1) *Iron Deficiency Experiment*: As discussed, plants under nutrient stress display a dynamic change in appearance and shape in leaves. For iron deficiency specifically, younger leaves turn yellow and change shape. This is difficult for static image processing methods. We observe our Grounded-SAM algorithm's ability to capture whole plant segmentation throughout time of an iron-deficiency experiment where four samples are under iron deficiency and four are given full nutrients. We found that our algorithm performed better throughout the entire experiment versus the original k-means segmentation method in the SMART pipeline (Fig 5). Most importantly, our algorithm is able to capture traits better at the end of the life cycle of the plant while the k-means algorithm becomes unstable. This will enable researchers to observe the late-stage effects of stress on traits more clearly.

2) *CVPR Leaf Counting Data Set*: To observe the ability of our algorithm to translate to other data sets without modification, we processed the CVPR leaf counting challenge data set [4]. Table I shows our algorithm performed well; but failed to achieve the performance of the algorithms developed for the challenge, including one utilizing k-means [11]. Fig 6 displays how our algorithm was able to perform well on data that was similar to the iron deficient data set that the prompt engineering was tuned for. But it fails to adapt to size outliers in the data set, such as a plant that is just a few pixels. On the other end, our algorithm adapt to the opposite end of size outlier with plants that are so large that the plant exceeds the image borders.

This experiment illustrated an important design consideration with integrating foundational model in vision. We support fellow literature on Segment-Anything having powerful zero-shot capabilities but requiring more exploration and fine-tuned

Method	FBD (A1)	FBD (A2)
Nottingham	95.3	93.0
MSU	94.0	87.7
Wageningen	94.7	95.1
EM + k-means	94.9	94.8
Grounded-SAM (ours)	86.6	82.8

TABLE I: Performance of Segmentation Algorithms in the CVPR Leaf Counting Challenge

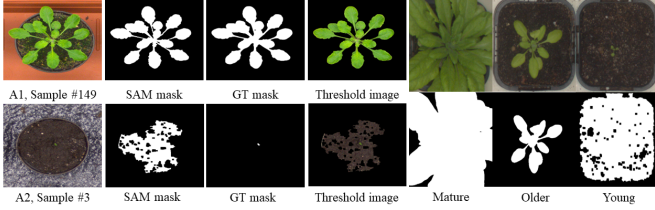


Fig. 6: Our SAM algorithm performed well on the majority of data samples but struggles with images that contain very small seedlings, which the fe-def data set contained no samples of.

data sets for real-life application areas, including agriculture [12].

B. Leaf-Specific Segmentation

Our experiment focused on the number of leaves detected and distribution of measurement values to illustrate the improvement in trait measurement performance. Fig 7 shows the improvement of the SAM algorithm over the SMART pipeline’s marker-based watershed algorithm. Our algorithm detected a more accurate increasing number of leaves per time point while the k-means and marker-based watershed started to fail at later time points. Furthermore, the distributions for trait measurements in area improved by having less unstable data samples. Another powerful effect of the SAM-based method is the high-resolution masks produced include other parts of the leaf including, the stem that was originally difficult to incorporate in trait analysis.

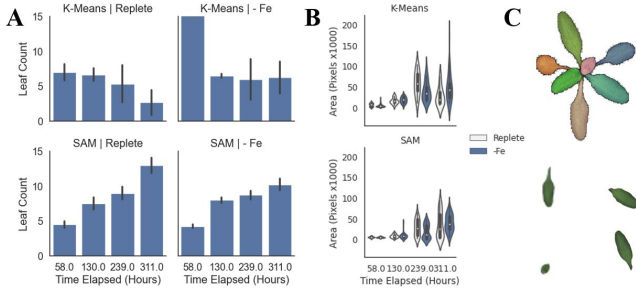


Fig. 7: (A) Our SAM algorithm had a strong ability to detect leaves through all stages of the plant’s lifetime. (B) The distribution of trait measurements in area improved with few outliers and smaller distributions using the SAM algorithm. (C) The masks and segmentations produced by SAM included the entirety of the leaf, including previously finicky stems.

V. CONCLUSION

Segment-Anything is a powerful zero-shot model for segmentation can accelerate many research areas. In this study, we showed the ability of SAM to produce high-resolution mask predictions that improve both segmentations and trait measurements. However, SAM is not a one-size-fits-all approach and requires refinement or fine tuning for a specific application area. Our algorithm failed to translate to another data set with different size extremes. In the future, we hope to explore more deeply the ability for SAM to be translated to a field situation and leverage the automatic method of segmentation. Furthermore, we hope to extrapolate dynamic insights from the high-resolution of trait measurements available from the algorithm for improving crop production.

REFERENCES

- [1] M. Khan, N. A. Castro-Guerrero, S. A. McInturf, N. T. Nguyen, A. N. Dame, J. Wang, R. K. Biundbeutel, J. Trupti, S. S. Jurisson, D. A. Nusinow, and D. G. Mendoza-Cozatl, “Changes in iron availability in arabidopsis are rapidly sensed in the leaf vasculature and impaired sensing leads to opposite transcriptional programs in leaves and roots,” *Plant, Cell, and Environment*, 2018.
- [2] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, “Segment anything,” 2023.
- [3] J. Adams, Y. Qiu, Y. Xu, and J. C. Schnable, “Plant segmentation by supervised machine learning methods,” *The Plant Phenome Journal*, vol. 3, no. 1, p. e20001, 2020. [Online]. Available: <https://access.onlinelibrary.wiley.com/doi/abs/10.1002/ppj2.20001>
- [4] H. Scharr, M. Minervini, A. P. French, C. Klukas, D. M. Kramer, X. Liu, I. Luengo, J.-M. Pape, G. Polder, D. Vukadinovic, X. Yin, and S. A. Tsafaris, “Leaf segmentation in plant phenotyping: a collation study,” *Machine Vision and Applications*, vol. 27, no. 4, pp. 585–606, May 2016. [Online]. Available: <https://doi.org/10.1007/s00138-015-0737-3>
- [5] M. Ren and R. S. Zemel, “End-to-end instance segmentation with recurrent attention,” 2017.
- [6] L. Luo, X. Jiang, Y. Yang, E. R. A. Samy, M. Lefsrud, V. Hoyos-Villegas, and S. Sun, “Eff-3dpsg: 3d organ-level plant shoot segmentation using annotation-efficient deep learning,” *Plant Phenomics*, vol. 5, p. 0080, 2023. [Online]. Available: <https://spj.science.org/doi/abs/10.34133/plantphenomics.0080>
- [7] R. Zenkl, R. Timofte, N. Kirchessner, L. Roth, A. Hund, L. Van Gool, A. Walter, and H. Aasen, “Outdoor plant segmentation with deep learning for high-throughput field phenotyping on a diverse wheat dataset,” *Frontiers in Plant Science*, vol. 12, 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpls.2021.774068>
- [8] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” 2020.
- [9] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu *et al.*, “Grounding dino: Marrying dino with grounded pre-training for open-set object detection,” *arXiv preprint arXiv:2303.05499*, 2023.
- [10] A. Reinke, M. D. Tizabi, C. H. Sudre, M. Eisenmann, T. Rädtsch, M. Baumgartner, L. Acion, M. Antonelli, T. Arbel, S. Bakas, P. Bankhead, A. Benis, M. Blaschko, F. Büttner, M. J. Cardoso, J. Chen, V. Cheplygina, E. Christodoulou, B. Cimini, G. S. Collins, S. Engelhardt, K. Farahani, L. Ferrer, A. Galdran, B. van Ginneken, B. Glocker, P. Godau, R. Haase, F. Hamprecht, D. A. Hashimoto, D. Heckmann-Nötzel, P. Hirsch, M. M. Hoffman, M. Huisman, F. Isensee, P. Jannin, C. E. Kahn, D. Kainmueller, B. Kainz, A. Karagyris, A. Karthikesalingam, A. E. Kavur, H. Kennigott, J. Kleesiek, A. Kleppe, S. Kohler, F. Kofler, A. Kopp-Schneider, T. Kooi, M. Kozubek, A. Kreshuk, T. Kurc, B. A. Landman, G. Litjens, A. Madani, K. Maier-Hein, A. L. Martel, P. Mattson, E. Meijering, B. Menze, D. Moher, K. G. M. Moons, H. Müller, B. Nichyporuk, F. Nickel, M. A. Noyan, J. Petersen, G. Polat,

S. M. Rafelski, N. Rajpoot, M. Reyes, N. Rieke, M. Riegler, H. Rivaz, J. Saez-Rodriguez, C. I. Sánchez, J. Schroeter, A. Saha, M. A. Selver, L. Sharan, S. Shetty, M. van Smeden, B. Stieltjes, R. M. Summers, A. A. Taha, A. Tiulpin, S. A. Tsaftaris, B. V. Calster, G. Varoquaux, M. Wiesenfarth, Z. R. Yaniv, P. Jäger, and L. Maier-Hein, "Common limitations of image processing metrics: A picture story," 2023.

- [11] N. Al-Shakarji, Y. Kassim, and K. Palaniappan., "Unsupervised learning method for plant and leaf segmentation," Apr 2017, pp. 1–4. [Online]. Available: <https://ieeexplore.ieee.org/document/8457935>
- [12] W. Ji, J. Li, Q. Bi, T. Liu, W. Li, and L. Cheng, "Segment anything is not always perfect: An investigation of sam on different real-world applications," 2023.