RootSAM: Adapting SAM2.1 for Root Segmentation in Minirhizotron Imagery

Kinjalk Parth[1] Sebastian Varela[1] Andrew D.B. Leakey[1] Samuel Walker[1] Nadia Shakoor[2]
[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[2]Donald Danforth Plant Science Center, Olivette, MO, USA

kinjalk2@illinois.edu, sv79@illinois.edu, leakey@illinois.edu
samuel60@illinois.edu, nshakoor@danforthcenter.org

Abstract

Root analysis is fundamental to understanding plant adaptation to environmental stresses and optimizing agricultural productivity [2, 4]. Generalized training of foundation models such as Segment Anything Model (SAM) 2.1 [6] fail to capture the unique challenges of root imagery. We present RootSAM, a specialized adaptation of SAM2.1 that achieves 66% mean Intersection over Union (mIoU) and 85% reduction in manual annotation time through a novel two-stage fine-tuning strategy. This framework enables effective adaptation of foundation models for plant phenotyping and is expected to accelerate root research.

1. Introduction

Minirhizotron imaging enables non-destructive quantification of root system depth and distribution in field-grown crops [8]. Manual annotations of quantitative traits from these images is notoriously laborious, limiting the scale of phenotyping studies [2, 9]. Foundation models like Segment Anything Model (SAM) [3] and SAM2.1 [6] have demonstrated remarkable zero-shot, promptable segmentation capabilities. However, their performance can degrade in specialized domains, such as fine and low-contrast structures of plant roots against a complex soil background [7, 9]. We first show that the original SAM2.1 model cannot effectively segment root structures [1]. We then propose a novel two-stage sequential fine-tuning approach that efficiently adapts SAM2.1 for bridging this performance gap. Our results show that fine-tuning is essential for effectively applying foundation models to root segmentation. We finally produce RootSAM after the two stages of fine-tuning.

2. Materials and Methods

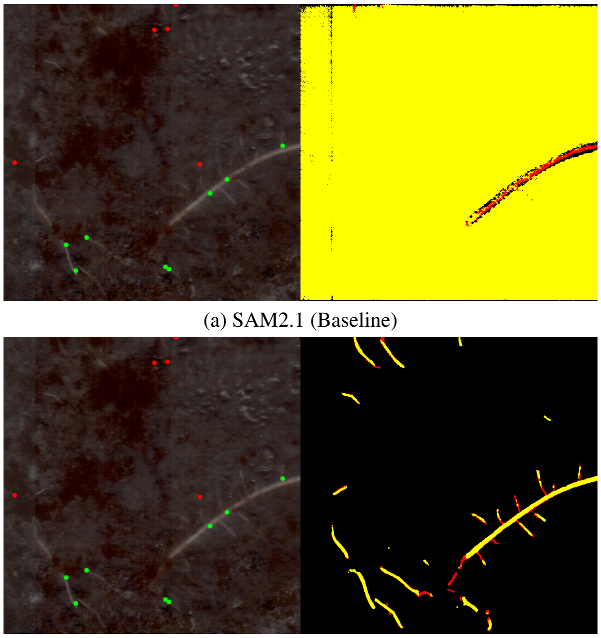
2.1. Dataset

Our dataset consists of 286 high-resolution color images of plant roots acquired using a CID Bio-Science CI-600 [1] in situ root imager. The device captures 360-degree scans from within minirhizotron tubes inserted into the soil. The resulting images present several challenges for automated segmentation, including variable illumination, and low contrast between roots and soil [9]. We divide this dataset into 228, 28, and 30 for train, validation, and test set.

2.2. Model Fine-Tuning

We utilized the SAM2.1 model with a Hiera-Large backbone as our foundation model [6]. To prepare the model for fine-tuning, we developed a prompt simulation strategy to mimic realistic user behavior, where prompts can occur anywhere in an object, not just its center [10]. Instead of morphological erosion, for each root mask, we generated positive (foreground) prompts by sampling points based on a probability distribution derived from a Euclidean Distance Transform [5]. This ensures the model is robustly trained on varied prompt locations. The fine-tuning process was then conducted in two stages, training only the decoder and prompt encoder to mitigate catastrophic forgetting and prevent overfitting. We focused exclusively on point-based prompts.

1. **Stage 1:** Images were upscaled to 2560×2560 px from their native resolution of 2550×2160 px, and then divided into patches of 512×512 px. This stage tuned the decoder to recognize root-specific traits and soil patterns and reduce the influence of global relationships.
2. **Stage 2:** Images were upscaled to 2048×2048 px and divided into patches of 1024×1024 px, used to refine the model's understanding of root system connectivity and branching patterns.



(a) SAM2.1 (Baseline)

(b) RootSAM

Figure 1. Qualitative comparison of segmentation performance. The baseline SAM2.1 model (a) struggles with the specialized task of root segmentation, such as incorrectly segmenting the soil background as root structures. After fine-tuning, RootSAM model (b) accurately identifies the root area.

3. Results

Our two-stage fine-tuning approach demonstrates progressive improvement in root segmentation performance across different scales. We evaluate the model exclusively on 512×512 pixel patches (primary evaluation) and 1024×1024 pixel patches (cross-scale validation) to assess the framework's ability to learn hierarchical features without catastrophic forgetting. We present our results at two levels of granularity to evaluate which resolution most effectively supports our current annotation workflows for generating additional training data.

Primary Evaluation (512×512 patches): RootSAM achieves a mIoU of **66%** and Dice Score of **77.2%**, significantly outperforming the baseline SAM2.1 model (48.2% mIoU, 56.5% Dice Score) (Table 1). Training baseline SAM2.1 on Stage 1 alone yields 65.3% mIoU, demonstrating that our two-stage approach provides an additional 1.1% improvement while maintaining robust performance.

Cross-Scale Validation (1024×1024 patches): RootSAM achieves 49.5% mIoU on larger patches (Table 1), substantially outperforming the baseline SAM2.1 which achieves only 16.3% mIoU, a remarkable 204% improvement. Baseline SAM2.1 with Stage 1 training alone achieves

Table 1. Performance comparison demonstrating progressive improvement through our two-stage strategy.

Model Configuration	mIoU	Dice Score	mAP
Primary Evaluation			
SAM2.1 (Baseline)	0.482	0.565	0.636
SAM2.1 (Stage 1)	0.653	0.764	0.803
RootSAM (Stage 1 + Stage 2)	**0.660**	**0.772**	**0.811**
Cross-Scale Validation			
SAM2.1 (Baseline)	0.163	0.211	0.226
SAM2.1 (Stage 1)	0.474	0.619	0.654
RootSAM (Stage 1 + Stage 2)	**0.495**	**0.639**	**0.667**

47.4% mIoU on larger patches, while our complete two-stage approach improves performance by 4.4% from Stage 1, therefore the full model demonstrates that it now incorporates global context learning without forgetting the fine-grained features learned in Stage 1.

Cross-scale validation provides a relevant contribution to fine root detection, where contextual information helps resolve ambiguities between roots and organic debris while maintaining the ability to capture intricate root structures. This progressive learning framework ensures that the model enhances its capabilities across scales rather than simply trading off between fine-grained and global understanding.

The model demonstrates robust performance, despite the constraints imposed by fine-tuning on a limited dataset. In comparison, other fully automated CNNs have reported lower scores even on images from more controlled environments [9]. This demonstrates the effectiveness of our fine-tuning strategy in upgrading a foundational model for plant root segmentation.

4. Conclusion and Future Work

We have presented a framework for efficiently fine-tuning the SAM2.1 model for root segmentation using minirhizotron imagery. Our two-stage sequential fine-tuning strategy enables the rapid development of a specialized segmentation tool with minimal initial manual labeling. The resulting model achieves a mIoU of 66% on 512×512 patches, which can now be used to facilitate and accelerate our current root tracing workflows. This performance level provides immediate practical value for high-throughput phenotyping applications. While our end goal is to develop a model capable of segmenting even fine roots directly from native resolution CI-600 images, the current framework establishes a strong foundation for building more automated root phenotyping pipelines by significantly reducing human annotation effort.

References

[1] CID Bio-Science, Inc. CI-600 In-Situ Root Imager. https://cid-inc.com/plant-science-tools/leaf-area-measurement/ci-600-in-situ-root-imager/, 2024. Accessed: 2024-10-27. 1

[2] Grégoire T. Freschet, Loïc Pagès, Colleen M. Iversen, Louise H. Comas, Boris Rewald, Catherine Roumet, Jitka Klimešová, Marcin Zadworny, Hendrik Poorter, Johannes A. Postma, Thomas S. Adams, Agnieszka Bagniewska-Zadworna, A. Glyn Bengough, Elison B. Blancaflor, Ivano Brunner, Johannes H. C. Cornelissen, Eric Garnier, Arthur Gessler, Sarah E. Hobbie, Ina C. Meier, Liesje Mommer, Catherine Picon-Cochard, Laura Rose, Peter Ryser, Michael Scherer-Lorenzen, Nadejda A. Soudzilovskaia, Alexia Stokes, Tao Sun, Oscar J. Valverde-Barrantes, Monique Weemstra, Alexandra Weigelt, Nina Wurzburger, Larry M. York, Sarah A. Batterman, Moemy Gomes de Moraes, Štěpán Janeček, Hans Lambers, Verity Salmon, Nishanth Tharayil, and M. Luke McCormack. A starting guide to root ecology: strengthening ecological concepts and standardising root classification, sampling, processing and trait measurements. *New Phytologist*, 232(3):973–1122, 2021. 1

[3] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 1

[4] Alexander Lux and Thomas L. Rost. Plant root research: the past, the present and the future. *Annals of Botany*, 110(2): 201–204, 2012. 1

[5] C.R. Maurer, Rensheng Qi, and V. Raghavan. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270, 2003. 1

[6] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. 1

[7] Abraham George Smith, Eusun Han, Jens Petersen, Niels Alvin Faircloth Olsen, Christian Giese, Miriam Athmann, Dorte Bodin Dresbøll, and Kristian Thorup-Kristensen. Rootpainter: deep learning segmentation of biological images with corrective annotation. *New Phytologist*, 236(2): 774–791, 2022. 1

[8] Teofilo Vamerali, Marianna Bandiera, and Giuliano Mosca. Minirhizotrons in modern root studies. pages 341–361, 2012. 1

[9] Tao Wang, Mina Rostamza, Zhihang Song, Liangju Wang, G McNickle, Anjali S Iyer-Pascuzzi, Zhengjun Qiu, and Jian Jin. Segroot: A high throughput segmentation method for root image analysis. *Computers and electronics in agriculture*, 162:845–854, 2019. 1, 2

[10] Bin Xie, Hao Tang, Dawen Cai, Yan Yan, and Gady Agam. Self-prompt sam: Medical image segmentation via automatic prompt sam adaptation. *arXiv preprint arXiv:2502.00630*, 2025. 1