American Society of Agricultural and Biological Engineers
2950 Niles Road | St. Joseph MI 49085-9659 | USA
269.429.0300 | fax 269.429.3852 | hq@asabe.org | www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.202401097
Paper Number: 2401097

3D segmentation within the root system architecture using Point Transformer

Xuehai Zhou[1], Leshang Bai[1], Rui Xu[2], Rui Kang[3], Davoud Torkamaneh[4], Shangpeng Sun[1]

[1]Department of Bioresource Engineering, McGill University, Montréal, Québec

[2]School of Computer Science, McGill University, Montréal, Québec

[3]Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing, China

[4]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec City, Québec, Canada

Written for presentation at the
2024 ASABE Annual International Meeting
Sponsored by ASABE
Anaheim, CA
July 28-31, 2024

ABSTRACT. *This study delves into the vital role of the primary root within the root system architecture (RSA) of Arabidopsis thaliana, a critical factor for plant growth and adaptation to environmental stress. Despite significant strides in root phenotyping, there remains a disparity in research depth between subterranean root structures and above-ground plant parts. This work seeks to bridge this gap by focusing on 3D root phenotypic analysis and developing a novel 3D segmentation method to identify primary roots within RSA. Leveraging a small point cloud dataset with 20 training, 5 validation, and 10 testing samples, and deploying Point Transformer series neural networks, the study established a comprehensive process from data annotation to model validation. The Point Transformer V2 (PTv2) model exhibited superior performance, achieving an average mean Intersection over Union (mIoU) of 0.77 and mean accuracy (mAcc) of 0.89, substantiating its robustness in high-fidelity RSA representation. In comparison, Point Transformer V3 (PTv3) and Sparse U-Net, although effective, showed variations in performance under complex root morphologies. The findings underscore the potential of PTv2 as a leading tool for RSA analysis, pivotal for enhancing our understanding of plant physiology and aiding crop improvement strategies.*

Keywords. *Root system architecture (RSA), primary root, point cloud segmentation, point transformers.*

Introduction

The intricate architecture of plant root systems, pivotal to growth and development, plays a central role in water and nutrient uptake and metabolic exchange, with the primary root being a key component of this complex structure. Established during embryogenesis, the primary root harbors the root meristem, a crucial origin point for all subsequent root development

(Jürgens, 2001). The dynamic structure of the primary root, along with the lateral roots, each responds distinctively to environmental stimuli, underscoring the necessity for detailed analysis of their developmental characteristics and interactions over time (López-Bucio et al., 2003). Furthermore, understanding root system architecture (RSA) is essential not only for appreciating plant growth but also for enhancing crop productivity. The primary root's role within the RSA is particularly significant due to its responses to various stresses, which can be distinctly different from those of lateral roots (de Dorlodot et al., 2007). By examining these components in unison, we can begin to unravel the complex interplay between a plant's belowground structure and its aboveground vitality, as well as the interactions between roots and soil.

Arabidopsis thaliana, with its well-characterized root morphology, serves as an exemplary model for elucidating the complexities of RSA. Its relatively simple and accessible root system provides a clear window into understanding the broader implications of RSA in plant development and stress response, such as salinity (Duan et al., 2015). As a model organism, Arabidopsis has significantly contributed to our understanding of plant genetics and has been indispensable in genome analysis due to its small genome size, ease of genetic manipulation, and rapid life cycle (Meinke et al., 1998). However, capturing the full spectrum of root morphology requires advanced 3D root phenotyping, which often involves sophisticated imaging systems and data processing platforms, as exemplified by platforms like MultipleXLab (Lube et al., 2022). Although these methods offer rich, detailed data, they come with high costs and technical demands. Therefore, leveraging Arabidopsis as a starting point can pave the way for breakthroughs in 3D phenotyping innovation, making it more accessible for comprehensive root analysis in various plant species. By studying Arabidopsis, researchers can develop and refine 3D analysis methods that are scalable to more complex root systems, thereby advancing our understanding of RSA's role in plant physiology and its adaptive strategies under various environmental conditions.

The pioneering work by Clark et al. marked a significant leap forward in non-destructive 3D root phenotyping by using transparent growth cylinders filled with gellan gum, allowing for real-time analysis of root growth without disruption (Clark et al., 2011). This non-destructive approach opened new possibilities for studying root development over time. Liu et al. built upon these advancements and specifically targeted field-grown maize (Liu et al., 2021). They established an automated system for capturing multiview stereo images, enabling the high-throughput reconstruction of high-fidelity 3D root point clouds. Although the approach was destructive, it stood out for its speed and the quality of the 3D reconstructions it produced, representing a compromise between detail and throughput. The study by Zeng et al. brought a different perspective by utilizing X-ray computed tomography (CT) scans to analyze root systems (Zeng et al., 2021). This method, while more costly due to the specialized imaging required, provided unparalleled detail and a comprehensive understanding of root traits at both macroscopic and microscopic levels. Lastly, Wu et al. contributed to the field with a unique root imaging approach, employing a custom support mesh that allowed for non-destructive capture of root images (Wu et al., 2023). This method facilitated the repeated and accurate quantification of root architecture over the course of development. These studies each introduce unique advancements in root phenotyping, ranging from non-destructive analysis to high-throughput imaging. They utilize various imaging techniques to deepen our understanding of root system architecture, collectively demonstrating the field's progression towards more precise, efficient, and comprehensive methods for studying root growth and development.

Despite the significant advancements in root phenotyping detailed above, research focused on root systems—particularly beneath the soil—remains less explored compared to studies of above-ground plant parts such as shoots. Moreover, 3D analysis of plants is not as common as 2D approaches, highlighting a critical knowledge gap in the precise 3D phenotyping of RSA. The primary aim of this study is to develop a robust 3D segmentation method specifically tailored for identifying primary roots within the RSA. Our objectives are threefold: (1) to deploy 3D point cloud segmentation neural networks for detailed root analysis, (2) to establish an efficient protocol for annotating 3D root data, develop preprocessing techniques for these data, and train the segmentation models using our custom primary root dataset, and (3) to evaluate our method by comparing the outputs of our segmentation models against human-labeled data to verify accuracy and reliability.

Materials and Methods

Dataset. The dataset comprises 35 point cloud data of Arabidopsis thaliana roots derived from the work of Sultan et al.(Sultan et al., 2020), with two representative samples shown in Figure 1. It was selected from ten individual plants spanning various stages of growth, specifically around days 9, 11, 13, 15, 17, 21, and 25. Notably, the data from Sultan et al. contains only x, y, z information and is presented at different scales.

Method. Under the premise of having acquired high-precision point cloud data of root systems, the workflow for the 3D primary root segmentation task can be divided into several stages: data annotation, data preprocessing, data augmentation, and segmentation neural networks training (Figure 2).

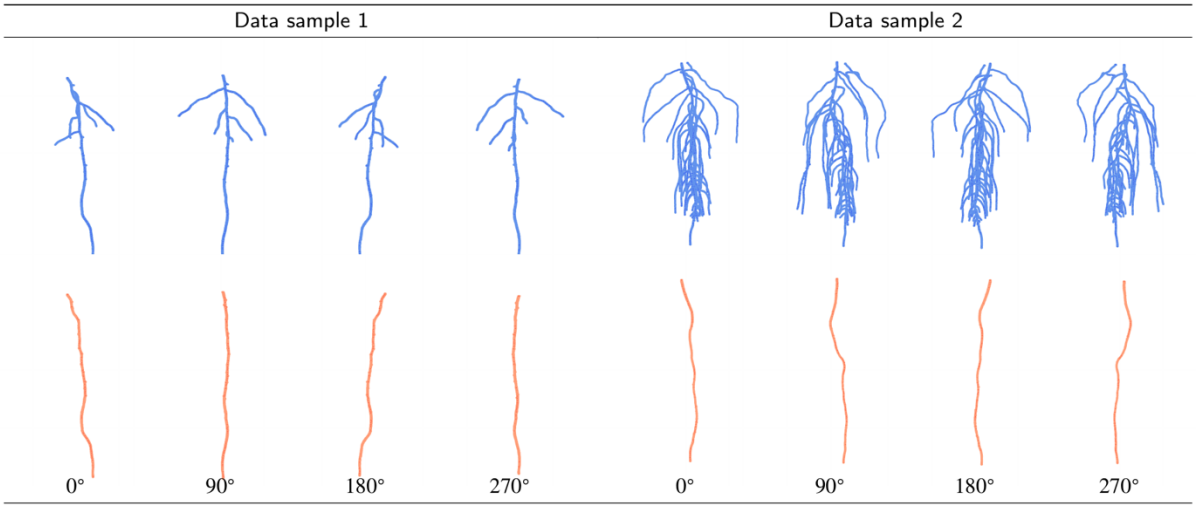

Figure 1: Multiview stereo visualization of original root point cloud and annotated primary root points. This figure presents two representative examples from the root dataset. The top row shows the original point cloud data, while the bottom row illustrates the annotated primary root.

Data annotation: In the context of 3D data manipulation, the current landscape of annotation tools exhibits a notable gap in sophistication compared to their 2D counterparts. This discrepancy is attributed to the relatively nascent stage of development in 3D methodological advances. Our evaluation focused on MeshLab and CloudCompare, which, despite not being expressly designed for 3D data annotation, are frequently repurposed for such tasks due to the dearth of specialized alternatives. A comparative analysis revealed that each platform presents distinct limitations in the context of 3D data annotation. MeshLab offers a point selection tool; however, it inherently selects clusters of points, which often leads to the inadvertent inclusion of superfluous data points, thereby compromising the precision of the annotation. Conversely, CloudCompare allows for point selection within 3D bounding boxes, which provides a more targeted approach to data curation. Nevertheless, it lacks an 'undo' feature, rendering any erroneous selections as significant setbacks, often necessitating a complete re-annotation of the dataset. Despite these challenges, CloudCompare was ultimately selected as our primary annotation tool. Even with its limitations, it presented a more manageable workflow for our purposes. Nevertheless, the manual annotation process for each root instance within the point cloud datasets was labor-intensive, averaging approximately 20 minutes per instance. This duration underscores the need for more efficient, purpose-built annotation tools in the field of 3D data processing to enhance productivity and accuracy.

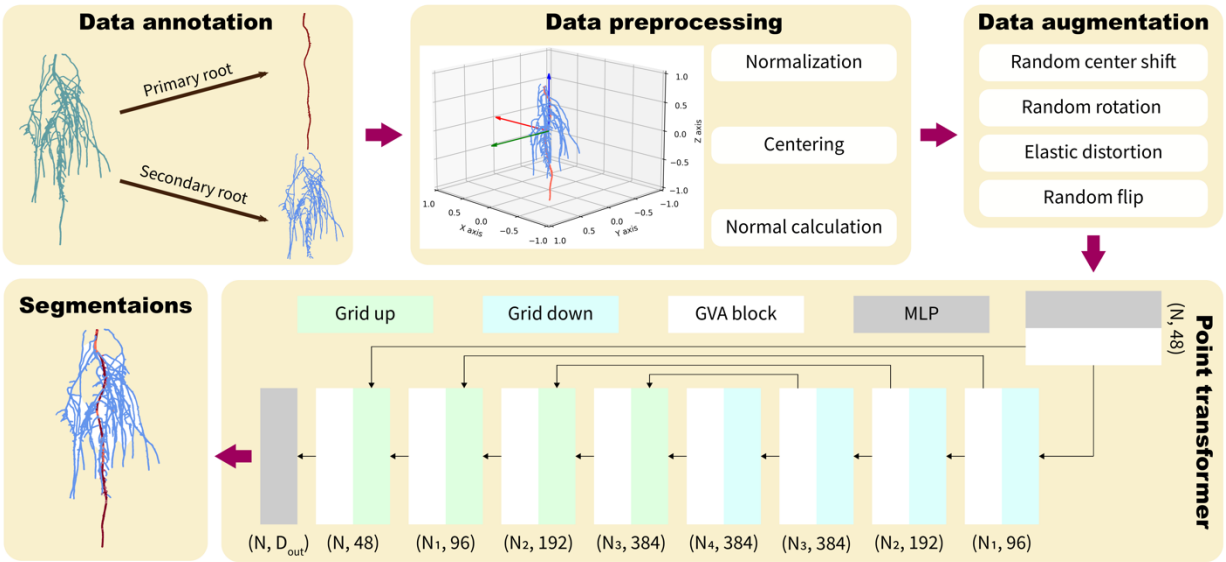

Figure 2: Workflow of 3D primary root segmentation. Initially, point clouds are annotated to identify primary and secondary roots. The annotated data is preprocessed through normalization to fit within a unit cube and centering at the origin, followed by the calculation of normals. Subsequently, augmentation techniques such random rotation are applied. The augmented data is then fed into point cloud segmentation neural networks for model training.

Data preprocessing: The dataset, once annotated, undergoes preprocessing which includes normalization to fit within a unit cube, followed by re-centering at the origin. Subsequently, normals are calculated and appended to point cloud data to enhance its representation for subsequent processing. The approach employs a KDTree, constructed from the 3D coordinates of the points, to expedite the search for nearest neighbors, which is crucial for the accurate estimation of local surface geometries. For each point in the cloud, the k nearest neighbors are identified, excluding the point itself, and the covariance matrix of these neighbors is computed to characterize the local surface structure. Singular value decomposition (SVD) is then applied to this covariance matrix to determine the principal components of the neighborhood distribution. The normal to the local surface is inferred from the last row of the matrix resulting from SVD, representing the component associated with the smallest singular value, and hence indicative of the least variance direction. To ensure that the normals are oriented outwardly, their direction is compared with the vector pointing from the centroid of the neighbors to the point in question. If a normal is found to be oriented towards the centroid, it is inverted to guarantee an outward orientation. This methodological approach ensures that the normals are not only accurately determined but are also consistently oriented relative to the surface they represent.

Data augmentation: Following the preprocessing of the annotated data, to fully exploit the potential of the original datasets, we employed a series of data augmentation techniques to increase their diversity. These augmentation methods include random center shift, random dropout, random rotation, random re-scaling, random flipping, chromatic transition, and random point sampling. These comprehensive augmentation techniques enable the neural network to learn an enhanced range of characteristics of the roots.

3D primary root segmentation models: Fully developed root systems are complex, with primary roots often developing in intricate, curly patterns to penetrate the soil effectively. Accurately encoding positional information is therefore crucial for neural networks to understand and learn the unique features and growth patterns of roots. Point Transformer (PT) series networks excels in this regard, making it an effective tool in agronomic research and related fields where precise and detailed segmentation of such complex biological structures is critical.

Point Transformer V2. Point Transformer V2 (PTv2) is a point cloud segmentation model that leverages the advantage of transformer-based architectures (Wu et al., 2022). It builds on the success of its predecessor Point Transformer (PTv1) (Zhao et al., 2021) by enhancing the ability to capture complex dependencies within point cloud data. The key innovations of PTv2 lie in the introduction of Grouped Vector Attention (GVA), Position Encoding Multiplier, and Partition-based Pooling.

- *Grouped Vector Attention (GVA):* GVA is an adaptation of the conventional attention mechanism, which was originally developed for handling scalar data such as text and image. In the context of point clouds, where data points are vectors, GVA modifies the attention mechanism to operate on vector data. This process involves three main steps: *grouping*, *vector attention*, and *aggregation*. Initially, points are grouped based on spatial or feature similarities using methods like clustering or neighborhood techniques. Within each group, vector attention is applied, enabling the model to capture detailed, group-specific features. The core of vector attention is calculated by eq (1)

$$\text{Vector Attention}(Q, K, V) = \text{softmax}(\frac{QK^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}) \odot V, \tag{1}$$

  where Q (Query), K (Key), and V (Value) are metrices derived from the input data, d_k is the dimensionality of the keys, and $\odot$ represents element-wise multiplication. This multiplication allows the attention scores to adaptively scale and rotate the vectors in the value matrix V, tailoring the output to the nuances of each group. Finally, the results from each group are combined through an aggregation step—such as summation or pooling—to produce the final output, effectively enhancing the model's ability to handle complex 3D structures in point cloud data.

- *Position Encoding Multiplier:* The position encoding multiplier significantly enhances spatial awareness and improves the flexibility and learnability of models dealing with point cloud data. This method encompasses three key steps: Positional features calculation, encoding and scaling, and multiplication with point features. Initially, for each point p_i in the point cloud, a positional feature vector f_i is calculated based on the spatial coordinates and potentially other local geometric attributes. This vector is then transformed through a learnable encoding, typically using a linear layer or a neural network, to map it to a higher-dimensional space, represented by

$$e_i = W \cdot f_i + b, \tag{2}$$

  where W and b are adjustable parameters. Differing from traditional methods that add positional encodings to input features, this approach integrates positional information by multiplying the encoded positional features e_i with the point features x_i, resulting in

$$x_i' = e_i \odot x_i. \tag{3}$$

This element-wise multiplication effectively embeds both the intrinsic and spatial properties directly into the feature representation, allowing the model to adaptively prioritize spatial relationships critical for interpreting and processing 3D environments.

- *Partition-based Pooling:* Partition-based pooling significantly enhances the model's ability to process complex spatial structures by preserving local details while integrating them into a global context. This pooling method consists of three main steps: partitioning the point cloud, aggregating features within each partition, and combining these features into a global descriptor. Initially, the point cloud is divided into partitions based on spatial or feature-based criteria, creating subsets P_i of points. Within each partition, a pooling operation, represented as

$$v_i = Pool(\{x_j | j \in P_i\}), \tag{4}$$

aggregates the features x_j of the points to produce a partition-specific feature vector v_i. Finally, these vectors are combined, often through concatenation or further processing, to form

$$v_{global} = Combine(v_1, v_2, \dots, v_n), \tag{5}$$

a comprehensive descriptor that encapsulates both local and global information of the point cloud, facilitating enhanced performance in tasks requiring detailed 3D structure interpretation.

We chose to explore 3D methods on RSA not only to achieve an intuitive expression of their morphology but also to take advantage of the morphological information to enhance model learning. PTv2 maximizes the potential of positional information using a position encoder and partition pooling, while also enhancing learnability by leveraging the Transformer architecture through GVA to discern the structural dynamics of RSA.

Point Transformer V3. Point Transformer V3, succeeding PTv2, represents the latest and current state-of-the-art in point cloud segmentation for extensive 3D scenarios such as autonomous driving. Inspired by the exceptional performance of large language models in Natural Language Processing (NLP) tasks, PTv3 seeks to enhance capabilities through various means, including scaling up the model size. The primary contributions of PTv3 include:

- *Emphasis on Scale:* PTv3 underscores the principle that scalability—in terms of handling larger datasets—plays a critical role in enhancing performance than the complexity of computational mechanisms. This focus on scale allows for broader applicability and improved generalization across diverse datasets.

- *Serialized Neighborhoods:* To overcome the computational overhead associated with traditional nearest neighbor search algorithms like K-Nearest Neighbors (KNN), PTv3 adopts a novel approach that utilizes pre-defined, serialized neighborhoods. These neighborhoods are organized based on specific patterns within the point clouds, enhancing processing efficiency and reducing computational costs.

- *Streamlined Architectural Design:* PTv3 simplifies its architecture by removing the dependency on relative positional encodings and integrating a pre-positive sparse convolutional layer. This modification significantly reduces memory requirements and accelerates processing times, facilitating faster model training and inference.

- *Expanded Receptive Field:* The efficient data processing capabilities of PTv3 enable it to consider a substantially larger receptive field around each point—up to 1024 points compared to just 16 in previous iterations. This expansion allows for a deeper and more comprehensive understanding of the 3D scene, improving the model's accuracy and robustness in spatial interpretations.

As a result, compared to its predecessor, PTv3 offers enhanced potential for establishing a more comprehensive primary root segmentation tool across various plant species, particularly when applied to large volumes of well-annotated data. This advancement can be adapted and performed reliably across diverse and challenging agricultural scenarios, thus setting a new benchmark in the precision and scalability of point cloud segmentation technologies in plant research.

Sparse U-Net. The Sparse U-Net used in this study serves as a benchmark network for processing point cloud data (Choy et al., 2019). The input point cloud is fed through a series of sparse convolutional blocks that perform feature encoding. These blocks utilize sparse convolutions to efficiently extract geometric features while managing the data's inherent sparsity, enabling downsampling that reduces resolution and allows the network to capture larger-scale features. During the decoding, or deconvolution stage, skip connections are employed to concatenate feature maps from the encoding path with their corresponding counterparts in the decoding path. This technique helps in recovering fine-grained details that may be lost during downsampling, thereby enhancing the model's ability to reconstruct detailed outputs.

Experimental platform. All experiments were conducted on a workstation featuring a MSI SUPRIM Liquid GeForce RTX 4090 graphics card, equipped with 24 GB of memory. The central processing unit (CPU) utilized was an Intel Core i5-13600K. The workstation operated under the Ubuntu 22.04 LTS operating system, which served as the platform for algorithm development and validation. The software environment included Python version 3.8. Deep learning models and related tasks were implemented using PyTorch version 1.13.1 and its accompanying library, torchvision, version 0.14.1. Additionally, image processing tasks were executed using the OpenCV library, version 4.7.0. Support for parallel computation was provided by the NVIDIA CUDA toolkit, version 12.0.

Evaluation metrics. The performance of the primary root segmentation networks is evaluated using two standard metrics: Mean Intersection over Union (mIoU) and Mean Accuracy (mAcc).

- *Average Mean Accuracy (Avg mAcc): Mean* Accuracy assesses the correct classification of the points for a single root instance of a point cloud. For a single root point cloud, it is calculated as

$$mAcc = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i+FN_i)}, \tag{6}$$

  where TP_i represents the true positives for class i, and FN_i represents the false negatives for class i, across all classes C within that single root point cloud instance. The mAcc is then the average of the individual Accuracies across all samples in the dataset

$$Average\ mAcc = \frac{1}{N}\sum_{j=1}^{N} mAcc_j, \tag{7}$$

  where N is the number of sample root instances in the validation set, and mAcc_j is the mAcc for the jth sample.

- *Average Mean Intersection over Union (Avg mIoU):* For a single root instance sample, mIoU for class i is computed as

$$mIoU_{class_i} = \frac{TP_i}{TP_i+FP_i+FN_i}, \tag{8}$$

  where TP_i is the number of true positives, FP_i is the false positives, and FN_i is the false negatives for class i. The average mIoU is the score for each class across all samples

$$Average\ mIoU = \frac{1}{N}\sum_{j=1}^{N}(C\sum_{i=1}^{C} mIoU_{i_j}), \tag{9}$$

  where $IoU_{class_{i_j}}$ is the mIoU for class i in individual root sample j, and N is the number of samples.

Results and Discussion

Networks training. The diversity in the convergence patterns underscores the unique learning mechanisms and potential trade-offs between complexity and performance across different neural network designs. We present the training dynamics of three neural network architectures, PTv2, PTv3, and Sparse U-Net, applied to our task of primary root segmentation in Figure 3. The training process was conducted over 100 epochs with a dataset comprising 20 training samples, 5 validation samples, and 10 test samples. The PTv2 model demonstrates rapid convergence to a lower training loss, stabilizing at approximately 0.1, indicative of efficient learning and model suitability for the primary root segmentation task. In contrast, although the PTv3 model maintains a general downward trend, it exhibits a higher degree of fluctuation in training loss, suggesting a potentially more complex learning pattern or sensitivity to the training data variations. Moreover, the Sparse U-Net architecture reveals a distinct behavior, converging at a higher loss level compared to PTv2. Despite the higher loss, the pattern of convergence is notably smooth, implying a consistent learning process across epochs. This pattern may reflect the inherent characteristics of the Sparse U-Net in processing the spatially sparse data, and its capability to capture the complex geometries inherent in root structures. Notably, all three models exhibit stable learning curves without evidence of overfitting, a testament to the robustness of the architectures and the effectiveness of the data augmentation strategies employed.

Table 1: Training and validation performance metrics for segmentation models. This table presents the training (with 20 samples) and validation (with 5 samples) performance for PTv2, PTv3, and Sparse U-Net. The performance is evaluated in terms of model efficiency, measured by parameter volume and training duration, as well as accuracy, assessed by Average mIoU and mAcc on the validation dataset. PTv2 demonstrates superior performance, achieving the highest Average mIoU with the shortest training time in our task.

Backbone network	Parameter volume	Training duration	Average mIoU	Average mAcc
PTv2	3908102	**0.18hr**	**0.78**	0.84
PTv3	46189838	0.37hr	0.77	0.85
Sparse U-Net	39156098	0.30hr	0.76	**0.87**

A summary of training and validation performance metrics for primary root segmentation models including PTv2, PTv3, and Sparse U-Net is provided in Table 1. The evaluation focuses on model efficiency, which is quantified through parameter volume and training duration, as well as precision, assessed by mIoU and mAcc on the validation dataset. From the presented data, PTv2 emerges as the most efficient model in terms of training duration, requiring only 0.18hr to complete the training process, which is much shorter than PTv3 and Sparse U-Net, with training durations of 0.37hr and 0.30hr, respectively. This efficiency is remarkable considering the complexity of the task and the size of the network indicated by its model size with a parameter volume of 3,908,102. In terms of validation accuracy, PTv2 again outperforms the others with a mIoU of 0.78, though it is closely followed by PTv3 with a mIoU of 0.77. Sparse U-Net, while demonstrating the highest mAcc of 0.87, exhibits a slightly lower mIoU of 0.7634, indicating a trade-off between general accuracy and class-specific intersection over union performance. These results highlight the trade-offs between speed, efficiency, and accuracy in model performance. While PTv2 shows superior performance in both training duration and mIoU, suggesting a high efficacy in segmenting the validation data, Sparse U-Net demonstrates a higher mean accuracy, potentially indicating a more consistent performance across primary and secondary root classes.

Figure 3: Training dynamics for root segmentation models. The dataset consists of 20 training samples, 5 validation samples, and 10 test samples. During training, the augmented data is fed into three distinct neural network architectures: PTv2, PTv3, and Sparse U-Net, for a total of 100 epochs. The training loss trends indicate stable learning without evidence of overfitting. Specifically, PTv2 achieves a convergence at a loss of approximately 0.1, whereas PTv3 and Sparse U-Net converge at a higher loss of approximately 0.2.

Segmentation results visualization. A visual comparison of the segmentation results achieved by PTv2, PTv3, and Sparse U-Net on two test samples representing distinct developmental stages of root structures is shown in Figure 4. The accuracy of the segmentation is color-coded: blue indicates the correct identification of primary root points, red signifies points of secondary roots erroneously classified as primary, and green denotes primary root points incorrectly marked as secondary. The test samples were selected to highlight the segmentation models' performance on varying complexities within point cloud data. For test sample 1, which illustrates an early-stage growth with a simple, straight elongated primary root, the models' ability to identify and differentiate the primary root points can be observed. On the other hand, test sample 2 showcases a more mature root system featuring a curvy primary root and an intricate network of secondary roots, posing a greater challenge for the segmentation models. The results for this sample provide insight into each model's capability to handle complex spatial structures and differentiate between primary and secondary root points under more challenging conditions.

Across both samples, PTv2 and PTv3 exhibit a higher degree of precision in identifying the main root structures with minimal misclassification, as evidenced by the prevalence of blue points and the sparse occurrence of red and green errors. Sparse U-Net, while still accurate in the primary root identification, shows slightly more misclassified points, particularly in the complex arrangement of test sample 2, which may be attributed to its distinct approach to handling spatial data sparsity.

Comparative analysis of testing results. A scatter plot that compares the performance of three segmentation models across ten test cases in the task of root segmentation is presented in Figure 5. Two key performance metrics, mIoU and mAcc, are plotted along the x-axis and y-axis respectively. PTv2 demonstrated an average mIoU and mAcc of 0.77 and 0.89 respectively, reflecting higher values in both metrics and suggesting its superior segmentation precision. Notably, PTv2 also exhibits a consistent trend of clustering toward the upper right quadrant, denoting a reliable segmentation capability across various test scenarios. PTv3, while achieving commendable results, with an average mIoU and mAcc mirroring PTv2's at 0.75 and 0.89 respectively, shows a data point with the lowest precision in terms of both average mIoU and mAcc, indicating some instances of underperformance compared to PTv2. Sparse U-Net, with an average mIoU of 0.69 and mAcc of 0.85, shows a spread distribution of results which implies variability in its segmentation performance. The disparity in its average

scores compared to PTv2 and PTv3 may point to a less consistent performance of segmentation applications across the test cases. Overall, PTv2 distinguishes itself with higher precision and consistency, establishing it as the most effective model among those tested for the primary root segmentation task.

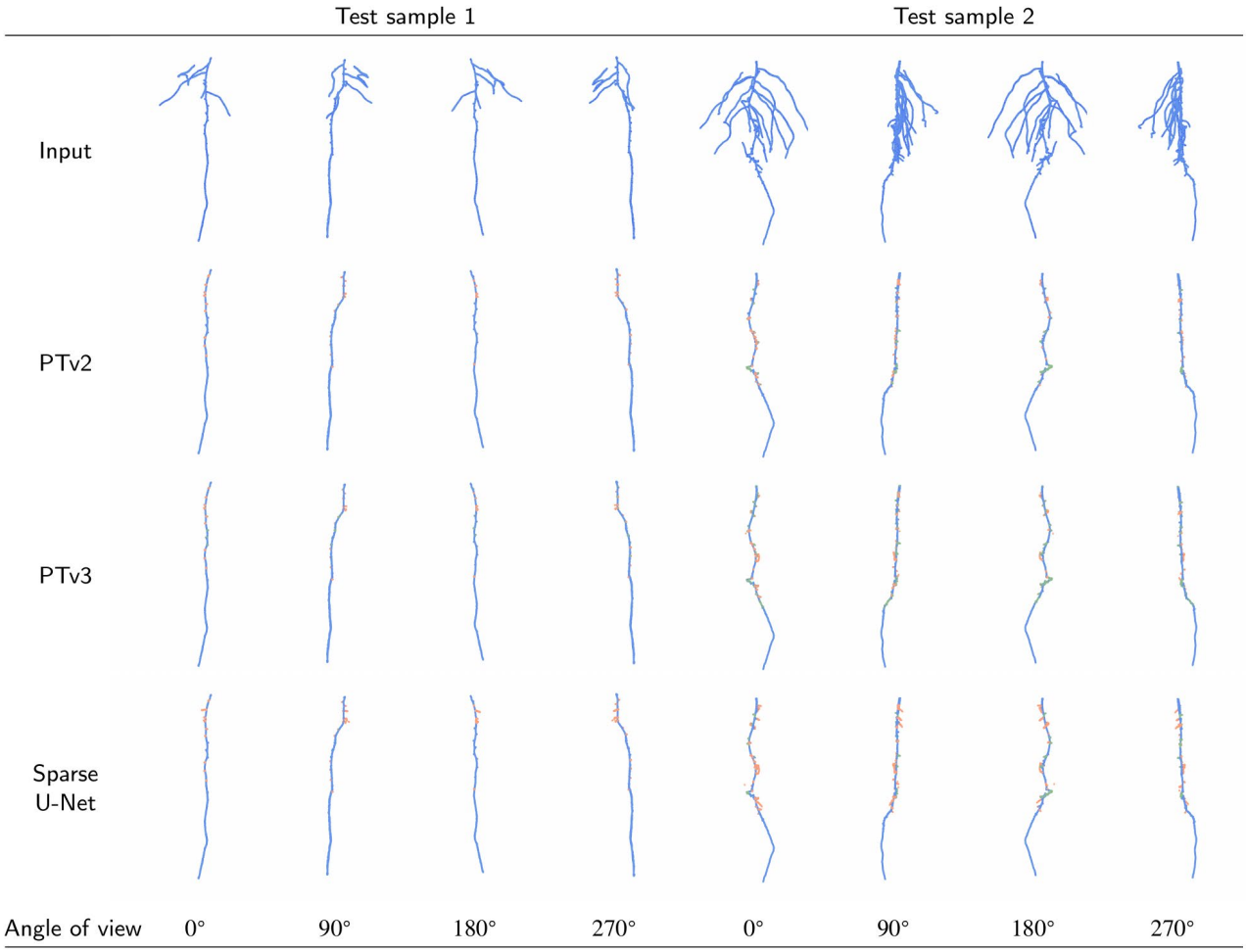

Figure 4: Comparative results of primary root segmentation on diverse test samples. Two testing samples representing different morphological and chronological stages are presented from multiple stereo angles. Sample one features an early-stage growth with a straight, elongated primary root. In contrast, sample two exhibits a later-stage development with a coiled primary root and complex secondary roots. Segmentation accuracy is color-coded: blue for correctly identified primary root points, red for secondary root points erroneously identified as primary, and green for primary root points mistakenly classified as secondary.

Discussion. Overall, across our primary root segmentation task, all three segmentation networks achieved plausible results. The PTv2 model, being significantly smaller than PTv3 and Sparse U-Net, required a correspondingly shorter training duration. Notably, with only 20 training samples, the compact PTv2 model exhibited marginally higher accuracy compared to the larger PTv3 and Sparse U-Net models. Although our segmentation networks performed well in identifying the main root in the majority of root instances, we observed that there is considerable room for improvement in identifying complex root structures, such as instances of spiraling root growth (Figure 6).

Contrary to our intuition, PTv3, which we hypothesized would surpass its predecessor PTv2, did not demonstrate superior performance in our task. We postulate that the primary reason for this outcome is that the large transformer architectures, like PTv3, which have an inherent predilection for larger datasets, while our training set consisted of only 20 samples. The labor-intensive and time-consuming nature of point cloud annotation means that high-quality annotated point cloud data has become a significant bottleneck in current scientific discovery. We believe that with access to abundant annotated data, models based on the PTv3 architecture have the potential to become universal primary root segmentation models across different species and varieties.

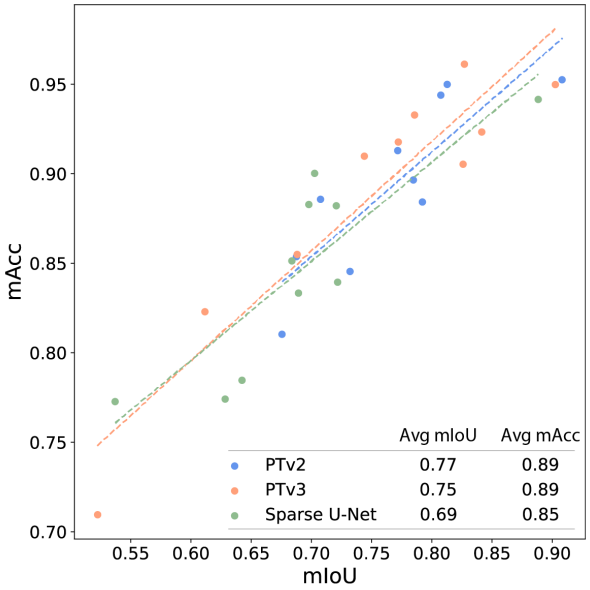

Figure 5: Comparative analysis of test results for root segmentation models. The scatter plot includes ten test results for PTv2, PTv3, and Sparse U-Net. Data points are plotted with mIoU on the x-axis and accuracy on the y-axis, position models with higher precision in the upper right quadrant. From the distribution, it is evident that PTv2 achieves superior precision among three evaluated models.

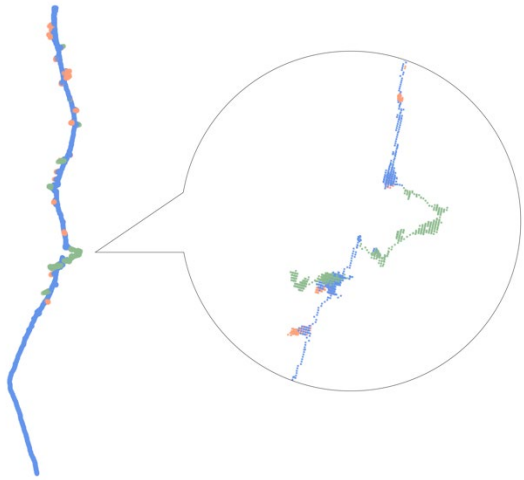

Figure 6: Representative segmentation result by PTv2 on test sample 2 at the 180° viewing angle. It is observable that in areas where the primary root is twisted, the segmentation network struggles to precisely classify those points correctly.

Future works. Moving forward, our research will focus on refining segmentation precision at the complex intersections of RSA, particularly in the challenging zones of spiraling growth. We will also expand our scope to assess the applicability of large-scale models for primary root segmentation across a wide range of plant species and varieties, aiming to create a more versatile and generalized tool. Furthermore, we intend to innovate in the representation of RSA, transitioning towards a skeletonized depiction that will facilitate a deeper morphological analysis. This consolidated effort is expected to enhance our understanding of root system dynamics and support the development of crops with optimized root traits for better environmental resilience.

Conclusions

In this study, we assessed the performance of three cutting-edge point cloud segmentation networks—PTv2, PTv3, and Sparse U-Net—on the challenging task of identifying primary roots from individual point cloud scans of Arabidopsis thaliana. Operating with a modest set of twenty training samples over 100 epochs, PTv2 emerged as the leading model, attaining the highest average mIoU at 0.77 and mAcc at 0.89 across an array of ten test datasets, while also demanding the shortest training duration within 0.18 hour. Though all three models generally demonstrated praiseworthy segmentation capabilities, their efficacy waned against the backdrop of complex root architectures, particularly those with twisting primary roots. Our future work is dedicated to enhancing the segmentation quality in such challenging scenarios, focusing on advanced algorithmic adaptations to handle the intricacies of root structure. Considering the crucial role of the primary root as an indicator of plant health and vitality, the application of the segmentation networks examined in this study is instrumental in the precise quantification of key traits such as root length and biomass, thereby enriching our understanding of RSA and its consequential effects on plant well-being and productivity.

References

Choy, C., Gwak, J., & Savarese, S. (2019). 4d spatio-temporal convnets: Minkowski convolutional neural networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Clark, R. T., MacCurdy, R. B., Jung, J. K., Shaff, J. E., McCouch, S. R., Aneshansley, D. J., & Kochian, L. V. (2011). Three-dimensional root phenotyping with a novel imaging and software platform. *Plant Physiology*, *156*(2), 455-465.

de Dorlodot, S., Forster, B., Pagès, L., Price, A., Tuberosa, R., & Draye, X. (2007). Root system architecture: opportunities and constraints for genetic improvement of crops. *Trends in plant science*, *12*(10), 474-481.

Duan, L., Sebastian, J., & Dinneny, J. R. (2015). Salt-stress regulation of root system growth and architecture in Arabidopsis seedlings. *Plant Cell Expansion: Methods and Protocols*, 105-122.

Jürgens, G. (2001). Apical–basal pattern formation in Arabidopsis embryogenesis. *The EMBO journal*.

Liu, S., Barrow, C. S., Hanlon, M., Lynch, J. P., & Bucksch, A. (2021). DIRT/3D: 3D root phenotyping for field-grown maize (Zea mays). *Plant Physiology*, *187*(2), 739-757. https://doi.org/10.1093/plphys/kiab311

López-Bucio, J., Cruz-Ramırez, A., & Herrera-Estrella, L. (2003). The role of nutrient availability in regulating root architecture. *Current opinion in plant biology*, *6*(3), 280-287.

Lube, V., Noyan, M. A., Przybysz, A., Salama, K., & Blilou, I. (2022). MultipleXLab: A high-throughput portable live-imaging root phenotyping platform using deep learning and computer vision. *Plant Methods*, *18*(1), 38.

Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). Arabidopsis thaliana: a model plant for genome analysis. *Science*, *282*(5389), 662-682.

Sultan, S., Snider, J., Conn, A., Li, M., Topp, C. N., & Navlakha, S. (2020). A statistical growth property of plant root architectures. *Plant Phenomics*.

Wu, Q., Wu, J., Hu, P., Zhang, W., Ma, Y., Yu, K., Guo, Y., Cao, J., Li, H., & Li, B. (2023). Quantification of the three-dimensional root system architecture using an automated rotating imaging system. *Plant Methods*, *19*(1), 11.

Wu, X., Lao, Y., Jiang, L., Liu, X., & Zhao, H. (2022). Point transformer v2: Grouped vector attention and partition-based pooling. *Advances in neural information processing systems*, *35*, 33330-33342.

Zeng, D., Li, M., Jiang, N., Ju, Y., Schreiber, H., Chambers, E., Letscher, D., Ju, T., & Topp, C. N. (2021). TopoRoot: a method for computing hierarchy and fine-grained traits of maize roots from 3D imaging. *Plant Methods*, *17*, 1-17.

Zhao, H., Jiang, L., Jia, J., Torr, P. H., & Koltun, V. (2021). Point transformer. Proceedings of the IEEE/CVF international conference on computer vision,