

# Root segmentation of horticultural plants in X-Ray CT images by integrating 2D instance segmentation with 3D point cloud clustering

Mary E. Cassity<sup>a</sup>, Paul C. Bartley<sup>b</sup>, Yin Bao<sup>a,c,\*</sup>

<sup>a</sup> Department of Biosystems Engineering, Auburn University, Auburn, AL, USA

<sup>b</sup> Department of Horticulture, Auburn University, Auburn, AL, USA

<sup>c</sup> Current address: Department of Plant and Soil Sciences, Department of Mechanical Engineering, University of Delaware, Newark, DE, USA



## ARTICLE INFO

### Keywords:

DBSCAN  
Instance segmentation  
Root segmentation  
X-Ray CT

## ABSTRACT

X-ray computed tomography (CT) is a powerful tool for in situ plant root system architecture (RSA) characterization. Accurate root segmentation from CT images is integral to studying RSA. Research studies on segmenting roots from CT images have been mainly limited to image processing-based approaches which may require parameter tuning and often lack common segmentation metrics, e.g., Dice and IoU. A recent deep learning approach utilizes a volumetric encoder-decoder network to achieve a high Dice score and IoU. However, training a volumetric model is dependent on the availability of fully annotated scans of the growing medium column, obtaining which can be time-consuming, tedious, and resource intensive. In this study, an efficient method using deep learning-based instance segmentation in conjunction with density-based spatial clustering of applications with noise (DBSCAN)-based filtering was developed and evaluated for two horticultural plant species. A pre-trained Mask R-CNN model was fine-tuned on images selected along different axes of the three-dimensional scans to identify the best view selection strategy for volumetric root segmentation. DBSCAN was used to filter noise from the volumetric segmentation with an automated parameter tuning technique. The proposed method was evaluated on scans of poinsettias and onions and achieved best average scores of 0.831, 0.839, 0.834, and 0.718 for Precision, Recall, Dice, and IoU, respectively. Further experiments showed reducing the training data to 1 % did not significantly impact the segmentation accuracy. Therefore, the proposed method has promising potential to facilitate RSA analysis with its high utility.

## 1. Introduction

Crop phenotyping is important for breeding resource-efficient crops, but conventional phenotyping methods are often limited to what can be seen above ground. However, root systems play a vital role in many plant functions, and a deeper understanding of how root systems interact with the environment can lead to the development of more resilient crops. Roots respond to both biotic and abiotic stress, reacting by changing their architecture in response to various conditions [3,25]. Moreover, root system architecture (RSA) determines a plant's distribution of resource capture and transport functions [13]. Therefore, quantitatively capturing RSA is an important part of plant phenotyping. Invasive approaches such as removing the root system from the soil have been used to phenotype RSA. However, these methods deform the root's three-dimensional (3D) structure and often cause a loss of root mass either through loss of fine roots during removal or after removal from

soil [14,30]. Other studies have indicated error in overestimating biomass measurements due to the adherence of non-root particles [11]. Furthermore, invasive methods do not allow for replanting the roots in the same 3D configuration they occupied before, making it impossible to analyze the root system's spatial growth over time [16]. Research in segmentation of roots grown in soil or a growing medium [12,24,32,34] or a gel medium [6,28] from two-dimensional (2D) images taken of cross sections of soil profiles is common. These methods offer a solution for low cost and rapid imaging of in situ root systems over time, however, they do not capture a 3D view of the root system.

X-ray computed tomography (CT), originally developed by G.N. Hounsfield for medical and clinical imaging [16] offers a tool for non-invasive RSA phenotyping for plants grown in containers without disturbing the roots and the soil or growing medium. Advances in CT technology that allow for higher resolution imaging have given X-ray CT a place in phenotyping RSA [16]. Furthermore, X-ray CT has been

\* Corresponding author.

E-mail address: [yinbao@udel.edu](mailto:yinbao@udel.edu) (Y. Bao).

evidenced as a tool to detect both biotic and abiotic stress [29]. A basic method for segmenting root systems from X-ray CT images is thresholding [7,23]. However, characteristics of root systems and the substrate they grow in pose significant challenges to accurate segmentation of root systems from X-ray CT images, including, overlap in attenuation values between root matter and water or other organic matter in the substrate and root topology such as negative geotropism roots [16]. Advanced segmentation methods leverage 2D and 3D image processing techniques to mitigate the inaccuracies introduced by overlapping attenuation values and diverse root topology. RooTrak [15] uses a top-down tracking approach with a level-set method [22,27] and updates based on local models of attenuation values. However, error can accumulate as the model progresses through the image stack, which is mitigated by only updating similar root shapes. Root1 [4] employs ImageJ tools for 3D segmentation, including normalization, edge detection, and median filtering, followed by bilevel thresholding and erosion to remove non-root materials, with user-assisted segmentation refinement using VGStudioMax. Rootine v.2 [17] enhances Rootine v.1 [8] by adding pot wall detection and background removal, using vesselness filtering [5] to differentiate true root structures, and incorporating post-processing to correct false negatives and remove false positives. RootForce builds on Rootine v.1 and uses vesselness filtering for fine roots and 3D Gaussian filtering for larger roots, effectively handling both fine and coarse root structures [9]. Each method offers specific advantages in handling root segmentation complexities, while their limitations include error accumulation, need for manual initialization, varying effectiveness across root sizes, and sensitivity to conditions of soil or growing medium.

In recent years, deep learning-based image segmentation methods showed impressive performance to handle variations in size, shape, and texture of X-ray root images through their hierarchical feature learning capability in large and diverse datasets. Deep learning models trained on volumetric datasets can utilize information from three dimensions to incorporate 3D shape, as opposed to two dimensions as commonly seen in segmentation tasks. Soltaninejad et al. [26] developed a deep learning-based parallel pipeline that combined both high- and low-resolution encoder-decoders with a multi-loss training approach for volumetric segmentation of roots from CT scans, outperforming Root1 and other deep learning-based semantic segmentation models. Selzner et al. [21] found that 3D U-Net segmentation to be an effective pre-processing step for magnetic resonance imaging of RSA, which significantly accelerated the subsequent manual root reconstruction, especially for cases of low contrast-to-noise ratios. One limitation to training a volumetric segmentation model is the need for large amounts of labeled 3D scans, which is more time-consuming, tedious, and resource intensive to obtain compared to their 2D counterparts. While semantic segmentation is commonly employed for segmentation of roots in CT images, it may suffer from the limitation that intersecting roots can form a single connected component in the segmentation mask, which causes challenges for the segmentation of individual roots in 3D space and introduces errors to morphological measurements of individual roots such as length, diameter, surface area, and volume. As roots tend to grow vertically, the cross-sectional regions of individual roots in CT slices along the vertical axis often appear in elliptical shapes with blob-like texture. Instance segmentation can potentially learn these visual features to better separate intersecting roots in each CT slice, facilitating tracking of individual roots across CT slices. To the best of our knowledge, the use of root instance segmentation has not been explored for root segmentation in X-ray CT images. To this end, the objective of this study was to develop and evaluate a 2D instance segmentation-based method for root segmentation in X-ray CT scans. We leveraged Mask R-CNN [10] and transfer learning to efficiently perform instance segmentation at the slice level for two container-grown horticulture crop species. The predicted 2D masks were merged into a volumetric segmentation of the whole root system. Furthermore, we evaluated different strategies of employing images sliced along different axes of the 3D scan and the efficacy of density-based spatial clustering of

applications with noise (DBSCAN) [18] as a post-processing step for improving volumetric segmentation accuracy. Lastly, we quantified the sensitivity of the root segmentation accuracy of our method as a result of reducing the amount of training images for instance segmentation.

## 2. Materials and methods

### 2.1. Image collection and annotation

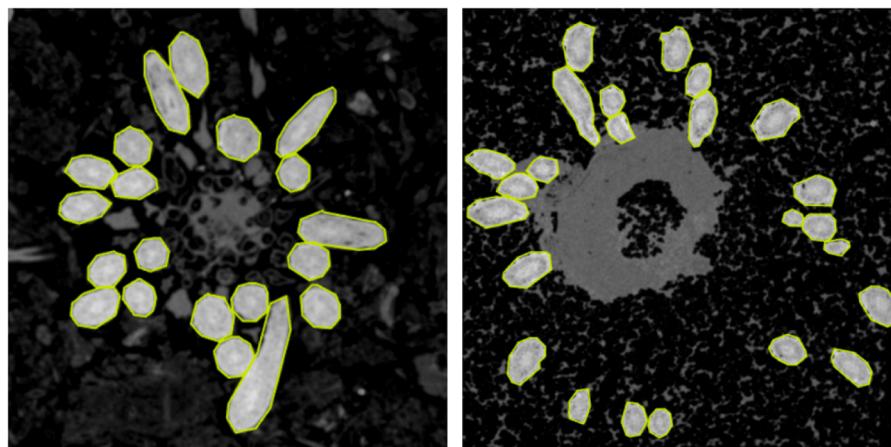
All plant root samples used in this study were scanned using a high-resolution X-ray CT scanner (XTH 225 ST, Nikon, Melville, NY) housed in the Shared Material and Instruments Facility at Duke University (Durham, NC). The instrument utilized a tungsten target, white beam, and conical beam with a detection panel containing  $2000 \times 2000$  pixels. A Feldkamp cone-based CT algorithm was applied to convert the X-ray radiographs into 2D reconstructed slices. A total of 2200 radiographs were collected from each scan by rotating the sample by  $0.164^\circ$  ( $2 \times$  frame averaging to reduce image noise). X-ray settings were adjusted to improve the image quality when scanning plant samples of differing volumes and moisture contents. Poinsettia (*Euphorbia pulcherrima* "Majestic Red") was asexually propagated and rooted in an engineered foam medium (87–50,010 RootCube Wedge; Oasis Grower Solutions, Kent, OH, USA). The poinsettia plants were scanned at 80 kV, 140  $\mu\text{A}$ . The two datasets selected from the poinsettia scans for further processing will be referred to herein as "Poinsettia1" and "Poinsettia2". Onion (*Allium cepa*) was grown in either a polyacrylic tube (6.35 cm internal diameter) or a custom peat-based sleeve. One onion sample was scanned at 95 kV, 120 mA, while the second onion sample was scanned at a higher moisture content at 110 kV, 120 mA. The two datasets selected from the onion scans for further processing will be referred to herein as "Onion1" and "Onion2". The images contain a voxel resolution of 30  $\mu\text{m}$  and 55  $\mu\text{m}$  for the poinsettia and onion, respectively. Ground truth annotations were obtained by manual annotation of root instances using COCO Annotator [2], resulting in a total of 8000 annotated images along the Z axis. Sample annotations are shown in Fig. 1. The annotated mask images were used to construct the ground truth of root reconstruction.

### 2.2. Data pre-processing

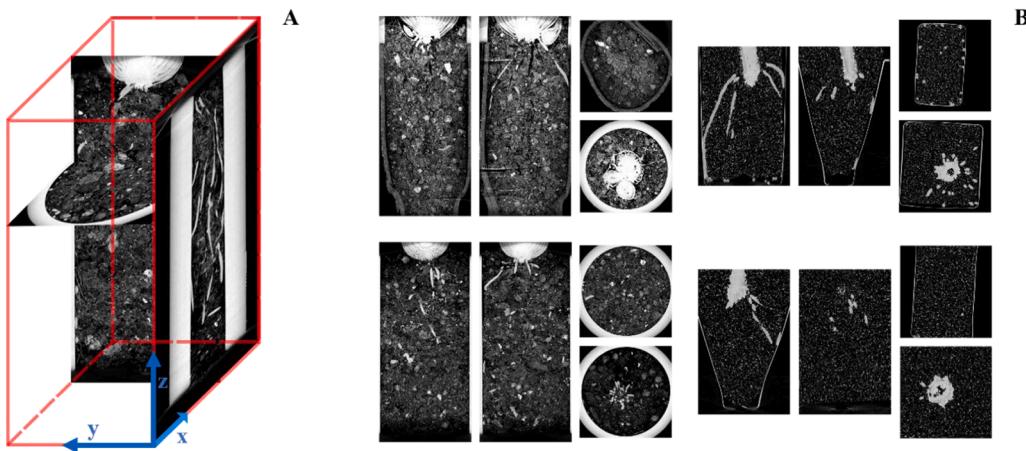
Two datasets were curated to assess whether fine-tuning a pretrained Mask R-CNN model on images from all three axes of the growing medium column and combining prediction results from multiple axes improved segmentation results. The first dataset was produced from the original X-ray CT images. These images were extracted along the Z axis of the growing medium column. The second dataset was produced by transforming both the input images and annotation masks from the first dataset. The images and masks were stacked sequentially to produce a volumetric image and mask for each plant. Next, these images were sliced along the X, Y, and Z axes of the volume to produce images and corresponding masks from each view of the growing medium column. Fig. 2 illustrates images from the X, Y, and Z axes and their orientation in the volumetric image and displays sample images from each plant.

In addition to using images from all axes of the growing medium column, data augmentation was performed by random cropping and rotation. Four tiles were randomly cropped from each input image and its corresponding ground truth mask and rotated randomly by  $0^\circ$ ,  $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $180^\circ$ , or  $270^\circ$ . The instances of roots in each mask were found using `findContours` function from OpenCV 4.7.0 [1] and the segmentations for each instance were saved in JSON files in the COCO format for model training.

With two scans for onion and poinsettia, respectively, two-fold cross validation was employed to robustly evaluate root segmentation accuracy. The four scans were split into two groups: Dataset1 (Onion1 and Poinsettia1) and Dataset2 (Onion2 and Poinsettia2). Furthermore, four total datasets were derived: Dataset1 Z, Dataset1 XYZ, Dataset2 Z, and



**Fig. 1.** Polygon annotations of individual roots along the vertical/Z axis.



**Fig. 2.** (A) Sample images and their orientations from the Onion2 growing medium column. Images are from slices along the X, Y, and Z axes. (B) Sample images from the Onion1 (top left), Poinsettia1 (top right), Onion2 (bottom left), and Poinsettia2 (bottom right) growing medium columns. For each image set, a sample image from the X axis (left), the Y axis (middle), and two images from the Z axis (right) of the growing medium column are displayed.

Dataset2 XYZ, where Z refers to datasets containing tiles from the Z view only and XYZ refers to datasets containing tiles from all X, Y, and Z views. The Mask R-CNN models trained on images from Onion1 and Poinsettia1 were evaluated on the images from Onion2 and Poinsettia2 and vice versa.

### 2.3. 2D root instance segmentation

Mask R-CNN models were trained using Meta AI Research's Detectron2 [33] and Python 3.10.9 [31]. The backbone and pretrained COCO Instance Segmentation model mask\_rcnn\_50\_FPN\_1x was used for training with a learning rate of 0.001, batch size of 64, and maximum iterations of 1000. The hyperparameters were selected by grid search. The default optimizer stochastic gradient descent (SGD) was used. Subsequently, the trained Mask R-CNN models were used to predict root instances in each scan. During prediction, each image was tiled with a 20 % side overlap. The predicted masks for each tile were recombined using a logical OR operation to make one predicted binary mask for each image. The models trained on images from the Z view were used to make instance predictions on images from the Z view. The models trained on images from XYZ view were used to make instance predictions on images from all three views. Sequential predicted masks were concatenated to produce a volumetric mask for the entire growing medium column. Seven different volumetric masks were made by combining the predictions on images from all three views, including: an X prediction, Y

prediction, Z prediction, XY prediction, XZ prediction, YZ prediction, and XYZ prediction. In addition, there were volumetric masks created by combining predictions made on the Z view by the model trained on the Z view. Therefore, eight total predictions existed for each plant.

### 2.4. 3D point cloud clustering and noise removal with DBSCAN

The proposed method aimed to incorporate 3D information in its methodology without needing to train a model using 3D images or relying on top-down or bottom-up connectivity. Therefore, Euclidean-based clustering for noise removal was explored and a filtering method using DBSCAN was implemented. Unlike most clustering methods, DBSCAN clusters data points based on density instead of distance, clustering continuous volumes of homogenous density in space similarly to how a human would. DBSCAN also accounts for continuity without using a top-down or bottom-up approach. Furthermore, DBSCAN is an unsupervised machine learning algorithm that can be implemented independent of the amount of training data used to train our instance segmentation model. Although DBSCAN is commonly used for noise removal, it can be difficult to select parameters to optimize noise removal for DBSCAN. Therefore, methodology was developed to automatically compute the input parameters for the DBSCAN algorithm and remove noise from the segmentation.

Before clustering our segmentation with DBSCAN, the dataset was downsampled to increase efficiency. Three dimensional NumPy arrays

consisting of a volumetric binary mask were made from the prediction segmentations saved in the JSON files. First, the segmentation contours were converted into a point cloud using Open3D 0.16.0 [35]. Open3D was used to downsample the data with voxel downsampling with a voxel size of 1.5.

DBSCAN requires two input parameters,  $\text{minPts}$  (the minimum number of neighbors a point must have to not be considered noise) and  $\epsilon$  (the radius in which DBSCAN can search for neighbors). Sander et al. [18] offered heuristics for choosing both parameters. It is common to use the  $\text{minPts} = 2 \times \text{dimensions}$ . For clustering in 3D space,  $\text{minPts}$  would equal six. This was experimentally found to be the optimal choice for  $\text{minPts}$  for the dataset through a grid search for  $\text{minPts}$ . Epsilon ( $\epsilon$ ) can be chosen by plotting the sorted k-nearest-neighbor distances computed for each point in the dataset [20]. This plot would have an “elbow” or “knee” point which indicates a value for  $\epsilon$ . This was computed using the Python package kneed after smoothing the k-nearest-neighbor distances plot with a moving average filter. The Python package was an implementation of the kneedle algorithm [19]. Any data point without the required number of minimum points in the radius  $\epsilon$  or not in the radius of a core point would be removed as noise.

Despite using the above methods for automatic parameter tuning, there was still noise present in the scans. The DBSCAN algorithm only removed noise if it belonged to a cluster consisting of six or fewer data points (voxels segmented as root). The predictions obtained through instance segmentation with Mask R-CNN contained noise belonging to clusters that consisted of more than six data points. Therefore, the technique described above to select  $\epsilon$  was adapted to filter possible noise clusters that were not removed by DBSCAN. The sorted number of points in each cluster were plotted and the knee point was computed using kneed after smoothing the plot with a moving average filter. This knee point indicates the point of maximum curvature on the plot of number of points in each cluster. It was assumed that noise clusters would contain significantly fewer data points than non-noise clusters, creating a clear point of maximum curvature. Clusters that were larger than the cluster size at the point of maximum curvature were kept while clusters that were smaller than the cluster size at the point of maximum curvature were removed as noise.

After noise was filtered from the segmentation, the point cloud was converted back into a 3D NumPy array as a volumetric binary mask. An empty 3D occupancy grid was initialized with the same resolution as the original CT scan, and the 3D coordinates of each point in the point cloud were mapped to the voxel indices. The state of the corresponding voxel was changed from 0 to 1. Next, dilation and erosion were applied to the

2D masks along the first axis using a  $5 \times 5$  kernel and one iteration in OpenCV. These operations were then repeated for the 2D masks along the third axis. This resultant upsampled volumetric segmentation were used to compute the accuracy metrics. Fig. 3 summarizes the prediction pipelines for both the Z axis and XYZ axes.

## 2.5. Evaluation

Precision, recall, dice (equivalent to  $F_1$ ), and IoU were used to evaluate the proposed segmentation and filtering method. These are commonly used metrics that describe both the successes (true positives and true negatives) and failures (false positives and false negatives) of the segmentation as well as overall segmentation accuracy. The definitions of these metrics are:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Dice} = \frac{2|G \cap S|}{|G| + |S|} \quad (3)$$

$$\text{IoU} = \frac{|G \cap S|}{|G \cup S|} \quad (4)$$

where

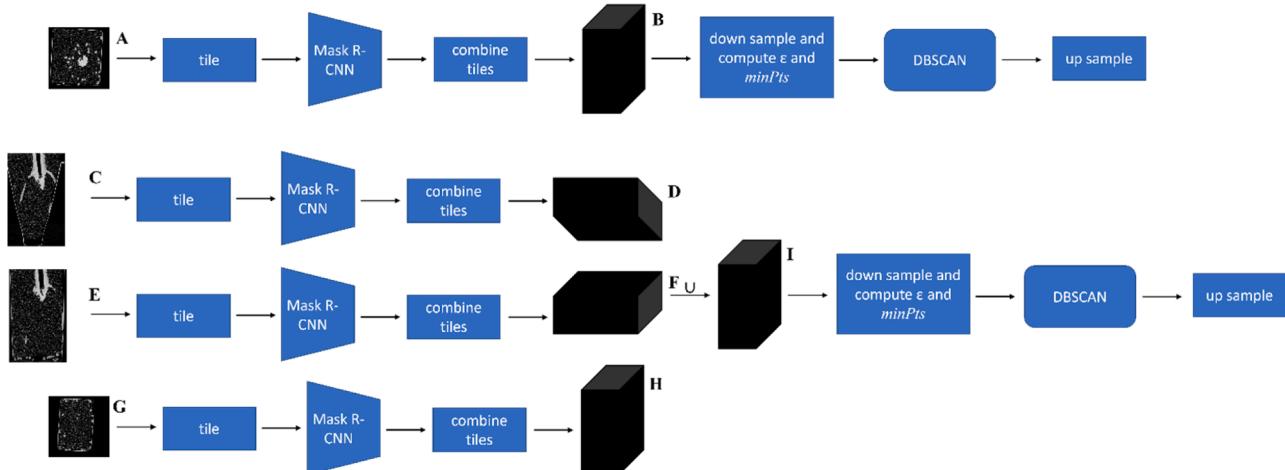
$\text{TP}$  = Number of correctly identified positive voxels (true positives),  
 $\text{FP}$  = Number of incorrectly identified positive voxels (false positives),

$\text{FN}$  = Number of incorrectly identified negative voxels (false negatives),

$G$  = Ground truth segmentation, and

$S$  = Predicted segmentation.

These metrics were computed for subsets of the entire scans. The subsets were manually selected to only include root material and represented the entire width and length of the original scans, but with truncated height (Z axis) to exclude non-root plant material at the top of the scans. Scores closer to 1 indicate better segmentation accuracy for all metrics. The metrics were computed by iterating through each voxel in the volumetric ground truth binary mask and the corresponding



**Fig. 3.** Segmentation method for the Z view (top) including: (A) input images from the Z view, (B) segmented volume from concatenating sequential Z predictions. Segmentation method for the XYZ views (bottom) including: (C) input images from the X view, (D) segmented volume from concatenating sequential X predictions, (E) input images from the Y view, (F) segmented volume from concatenating sequential Y predictions, (G) input images from the Z view, (H) segmented volume from concatenating sequential Z predictions, (I) segmented volume by combining multiple views.

prediction binary mask. To compare our method against a baseline approach, a global threshold was applied to the images of the growing medium columns. A grid search of the optimal threshold was performed in the range of 0 to 255 with a step size of 5.

## 2.6. Environment

The experiments were conducted on a laptop computer (Creator M16, MSI, Taiwan) with an Intel Core i7 CPU at 2.30 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU. Python 3.10.9 was used for the experiments.

## 3. Results

The evaluation metrics for root segmentation using Mask R-CNN only for the eight view selection strategies are summarized in [Table 1](#). The detailed results for each scan in each strategy are shown in Table S1 of the Supplementary Materials. The best strategy was training on the XYZ views and predicting on the Z view (Precision = 0.779, Dice = 0.826, IoU = 0.706), which improved the Precision, Dice, and IoU of the baseline strategy of both training and predicting on the Z view by 12.9 %, 5.8 %, and 9.0 %, respectively. On the other hand, the baseline strategy led to slightly higher recall than the best strategy (0.917 vs. 0.887). These differences in metrics suggested that the best strategy significantly reduced false positives, which outweighed the increase in false negatives compared to the baseline strategy. The two strategies of training on the XYZ views and predicting on either the XZ views or the YZ views performed closely and ranked between the best strategy and the baseline strategy in terms of all metrics. The other strategies were all trained on the XYZ views and performed worse than the baseline strategy. Among them, the strategy of both training and predicting on the XYZ views performed worst, which resulted in decreases in Precision, Dice, and IoU by 7.4 %, 3.3 %, and 5.9 %, respectively, and led to a slight increase in Recall by 2.1 %, with respect to the baseline strategy. These results indicate that training the model on the XYZ views acted as additional data augmentation that led to more accurate predictions on the Z view. As most of the roots grew vertically, the X and Y views tend to contain more elongated root images than the Z view that mostly contain circular cross sections of roots. As the X, Y, or XY views were included in addition to the Z view for prediction, the increased complexity of the unseen images might negatively impact the performance of the Mask R-CNN models.

Applying the DBSCAN algorithm to the Mask R-CNN-derived root volumetric segmentation slightly improved the Dice and IoU scores for all view selection strategies ([Table 1](#), the detailed results are shown in Table S2). Again, this was likely due to the reduction in false positives outweighing the increase in false negatives. In terms of Dice or IoU ranking, the strategy of both training and predicting on the XYZ views performed better than those of training on the XYZ views and predicting on either the X view or the Y view as a result of applying the DBSCAN-based filter. Even though the DBSCAN-based filtering did not improve the quantitative metrics by a large margin, it did significantly remove

many isolated small clusters that were false positives detected by the best Mask R-CNN model ([Fig. 4B](#) and C). From the qualitative results, the best Mask R-CNN model generally performed well on the two Poinsettia plants but occasionally failed to segment short lateral roots of small sizes (at the center of [Fig. 4B](#)) and the full length of some vertical roots ([Fig. 4B](#)) in Onion1. When the roots are dense and small in diameter as in Onion2, the best model tends to over-segment roots, which can be partially addressed by the 3D DBSCAN-based filtering ([Fig. 4D](#)).

In contrast, the thresholding-based segmentation produced far lower Precision, Dice and IoU scores than even the worst Mask R-CNN model ([Table 2](#)). Each scan resulted in different optimal ranges for thresholds, showing large variations in attenuation values between scans due to differences in root biomass, growing medium, moisture content, and scanning parameters. The qualitative results are not shown due to the extremely poor visibility of roots caused by the excessive false positives. The low accuracy metrics for the thresholding-based segmentation were caused by the overlapping attenuation values between non-root material in the growing medium column and the root material ([Fig. 5](#)). Qualitative comparisons of root segmentation between Mask R-CNN and the thresholding method are shown in Figures S1–S4 of the Supplementary Materials. The poor performance of the thresholding-based segmentation indicates that Mask R-CNN was able to leverage the morphological features of the cross-sectional regions of roots for improved segmentation.

[Table 3](#) shows the average Dice scores for all four scans using Mask R-CNN models with the DBSCAN-based filtering when training the models with different percentages of the full training data. No significant linear trend was observed as the training data were reduced for each view selection strategy. It suggests that there existed high redundancy in terms of root appearance in the complete training data. The fluctuations in mean Dice across percents of the full training data could be caused by differences in random sampling. The baseline view selection strategy of both training and prediction on the Z view had the second highest mean Dice score (0.820) using only 5 % of the full training data. Using the same amount of training data, the best view selection strategy of training on the XYZ views and predicting on the Z view achieved a slightly lower mean Dice score of 0.808 (i.e., a decrease of 1.5 %) than the baseline. Using 1 % of the training data, the baseline and the best strategies achieved the same mean Dice score of 0.794. This indicates that reducing the training data also reduced the benefit of data augmentation by using images along all three axes.

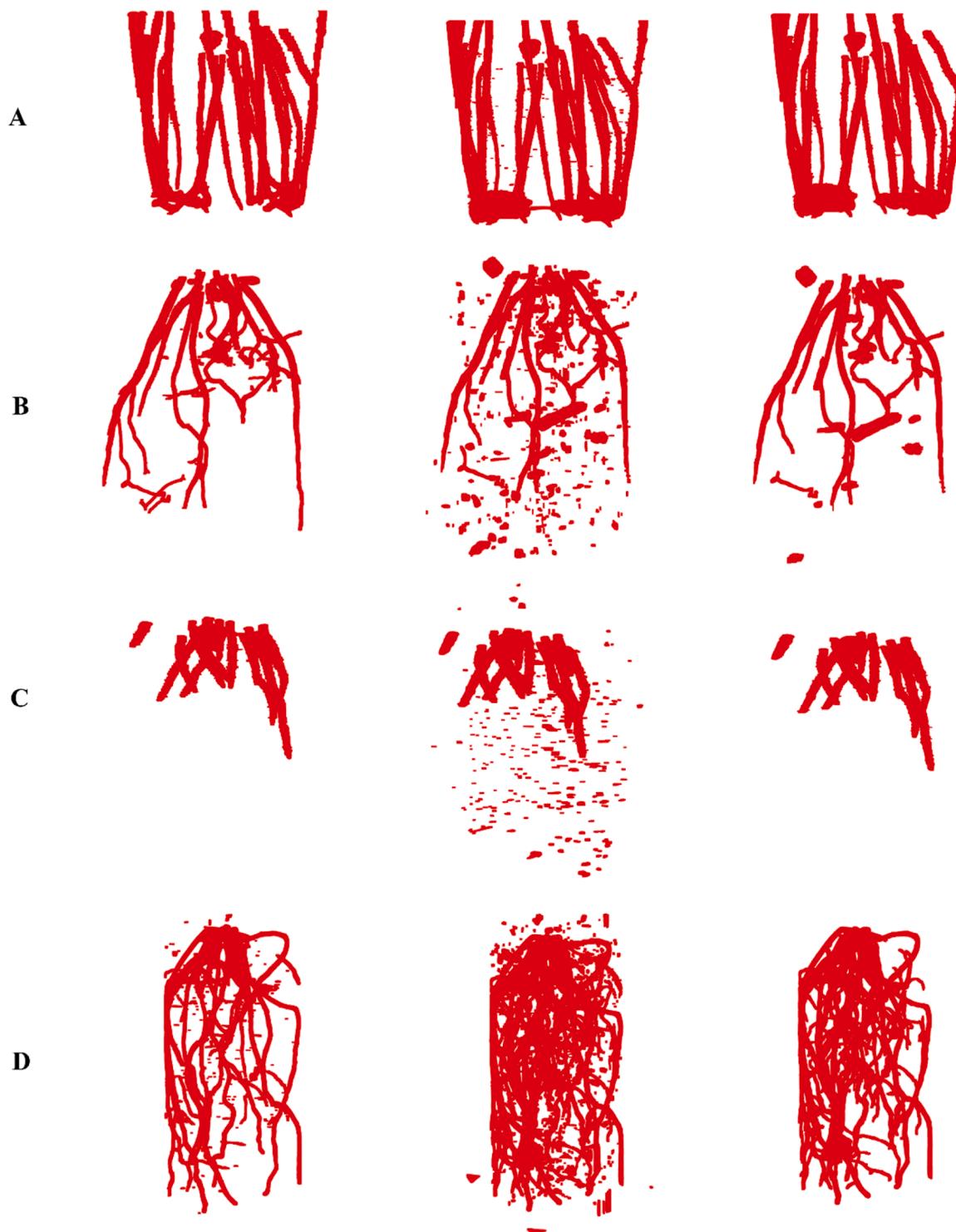
## 4. Discussion

The study presents an accurate and efficient method for root segmentation of horticultural plants in growing medium in X-ray CT images using 2D Mask R-CNN-based instance segmentation in conjunction with 3D DBSCAN-based filtering. The proposed method achieved high accuracy scores for Precision, Recall, Dice, and IoU, making it a promising approach for in situ RSA characterization. By utilizing 2D instance segmentation with transfer learning using a pretrained model, we capitalized on the benefits of automated feature learning by deep

**Table 1**

Average root segmentation accuracy metrics obtained by Mask R-CNN with and without 3D DBSCAN-based filtering for eight view selection strategies. The strategies are ranked based on Dice (IoU) without DBSCAN-based filtering in a decreasing order.

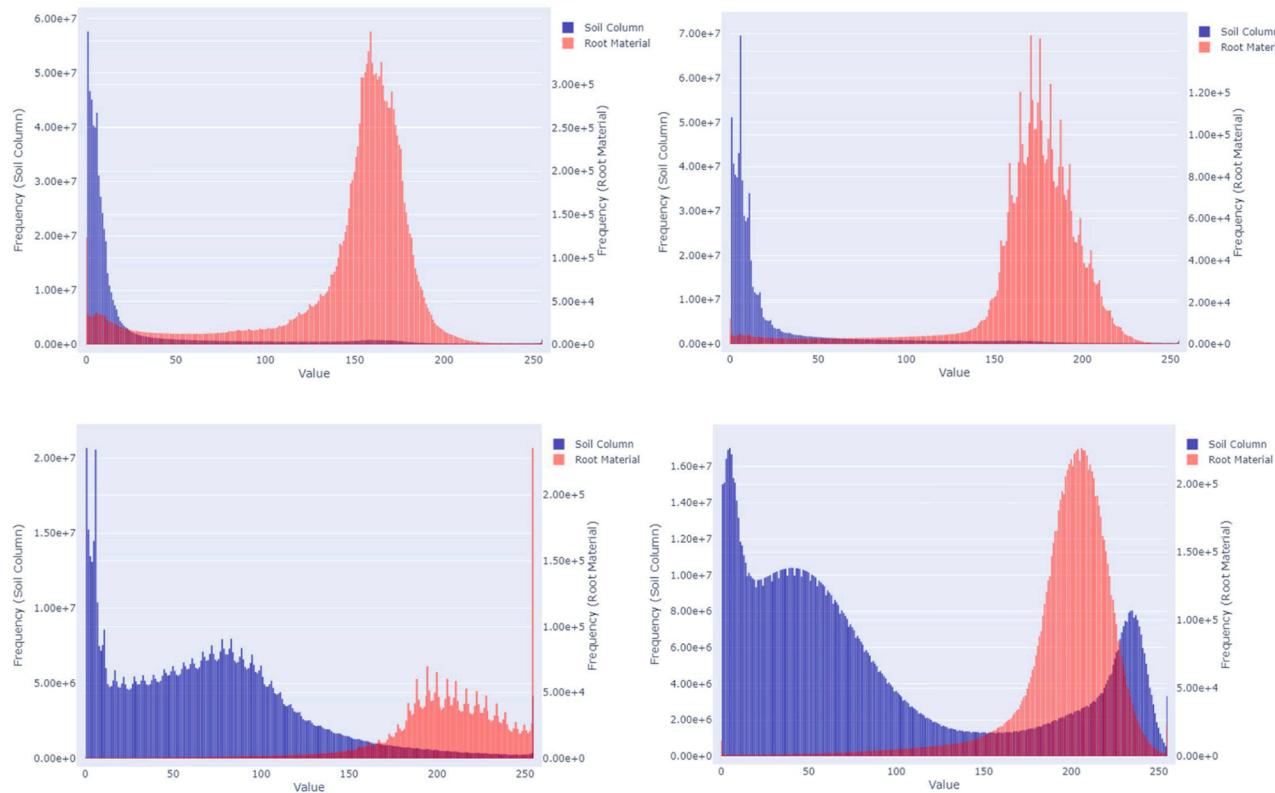
Training View(s)	Prediction View(s)	Precision		Recall		Dice		IoU	
		w/o	w/	w/o	w/	w/o	w/	w/o	w/
XYZ	Z	0.779	0.831	0.887	0.839	0.826	0.834	0.706	0.718
XYZ	XZ	0.692	0.729	0.923	0.913	0.787	0.808	0.652	0.681
XYZ	YZ	0.693	0.727	0.914	0.899	0.785	0.802	0.648	0.671
Z	Z	0.690	0.736	0.917	0.882	0.781	0.799	0.648	0.674
XYZ	XY	0.676	0.707	0.901	0.893	0.766	0.784	0.624	0.648
XYZ	X	0.757	0.671	0.792	0.926	0.761	0.776	0.620	0.636
XYZ	Y	0.736	0.806	0.791	0.756	0.754	0.770	0.612	0.633
XYZ	XYZ	0.639	0.787	0.936	0.739	0.755	0.754	0.610	0.615



**Fig. 4.** Ground truth (left), best Mask R-CNN model (middle), best Mask R-CNN model with DBSCAN-based filtering (right) for: (A) Poinsettia1, (B) Onion1, (C) Poinsettia2, (D) Onion2.

**Table 2**  
Segmentation accuracy metrics obtained by global thresholding.

Plant	Lower Threshold	Upper Threshold	Precision	Recall	Dice	IoU
Poinsettia1	145	195	0.412	0.701	0.519	0.350
Onion1	195	250	0.106	0.567	0.179	0.098
Poinsettia2	170	215	0.299	0.588	0.397	0.247
Onion2	185	210	0.086	0.451	0.145	0.078
Average	155	215	0.182	0.646	0.253	0.155



**Fig. 5.** Overlapped histograms of attenuation values between root and non-root voxels using thresholding-based segmentation for each of the four scans. Note that the two Y axes have different ranges and the bin with the highest frequency was removed for viewability.

**Table 3**

Average Dice scores of all four plants for training, prediction view(s) across percents of the full training data using Mask R-CNN and 3D DBSCAN-based filtering.

Percent of training data (%)	Z Z	XYZ X	XYZ Y	XYZ Z	XYZ XY	XYZ XZ	XYZ YZ	XYZ XYZ
1	0.794	0.707	0.657	0.794	0.620	0.691	0.642	0.604
5	<b>0.820</b>	0.780	0.737	0.808	0.708	0.752	0.725	0.694
10	0.819	<b>0.806</b>	0.774	0.818	0.762	0.779	0.772	0.744
15	0.792	0.787	0.704	0.813	0.679	0.762	0.695	0.666
20	0.806	0.683	0.682	0.703	0.621	0.638	0.627	0.600
25	0.803	0.768	<b>0.755</b>	0.814	0.714	0.742	0.739	0.698
50	0.802	0.615	0.658	0.741	0.573	0.596	0.645	0.559
75	0.806	0.759	0.671	0.831	0.641	0.750	0.676	0.635
100	0.799	0.770	0.754	<b>0.834</b>	<b>0.784</b>	<b>0.808</b>	<b>0.802</b>	<b>0.776</b>

convolutional neural networks for root segmentation. Furthermore, we leveraged the 3D volumetric nature of X-ray CT scans to extract slice images along all XYZ axes to increase the training data, which led to a boost in segmentation accuracy over using only the images along the Z axis. If resources for image annotation are limited, the pretrained Mask R-CNN models do not seem to require large amount of the labeled root images for fine-tuning to achieve good performance for the same plant species. Another notable aspect of the proposed method is the use of DBSCAN-based filtering with automated parameter tuning. Automated parameter tuning eliminated the need for human interaction that is commonly seen in image processing-based approaches. Furthermore, manual parameter tuning can be time-consuming and subjective. However, it should be noted that the noise removal implementation will always remove some of the root segmentation. For segmentations that do not contain a lot of noise (such as Poinsettia1) noise removal may end up removing root material resulting in a decrease in Dice or IoU. The user can forgo noise removal if the segmentation qualitatively appears to contain a small amount of noise. Alternatively, an automated check for whether further noise removal is needed can be implemented.

Regarding limitations, comparisons were not made between our method and existing methods due to the lack of publicly available software and datasets. For comparison with basic image processing-based approach, we showed that the thresholding-based segmentation failed to produce useful results for our dataset due to the overlapping attenuation values between root and non-root voxels. We speculate that complex image processing-based methods that rely on contrast between root and non-root voxels would suffer from the same challenge in our dataset. To compare with the 3D semantic segmentation-based method [26], we would need to train their model using our dataset because of large differences in root appearance and growing media between datasets. Since our dataset is far smaller compared to that of Soltaninejad et al. [26] (4 scans vs 47 scans), the 3D semantic segmentation model might not be able to train properly. It is worth mentioning that even annotating our small dataset required large amount of time and labor to reach satisfactory quality as the humans needed to trace roots across slices along different axes to have high confidence in the annotation. Our transfer learning strategy showed high utility for small datasets. Meanwhile, the strategy can potentially handle large diverse datasets by

fine-tuning pretrained instance segmentation models of deeper and more complex architecture.

It should be noted that instance segmentation in this study was primarily used to perform semantic segmentation of roots in CT slices. Whether instance segmentation models generally outperform semantic segmentation models for root segmentation remains uncertain. The instance identity information was not utilized in our current pipeline, as segmentation accuracy was evaluated at the root system level. Future research could explore developing a root skeletonization algorithm that traces root instance masks across slices based on multi-object tracking. This approach could enable individual root segmentation and facilitate morphological measurements such as length, branching, diameter, and biomass. In addition, a more comprehensive dataset including more plant species, growing media, moisture contents, and growth stages could be curated to compare our method with existing methods. A compilation of a large and diverse common dataset is integral to the development of segmentation methods and the comparison of segmentation accuracies across various methods. To facilitate such efforts, our dataset, models, and source code are made publicly available.

## 5. Conclusion

This study developed and evaluated a method for segmenting roots in X-ray CT scans of two horticultural plant species using 2D instance segmentation with Mask R-CNN and 3D DBSCAN-based filtering. Fine-tuning pretrained Mask R-CNN models on our dataset achieved satisfactory root segmentation accuracy given the challenge of the overlapping attenuation values between root and non-root voxels. The DBSCAN-based filtering was effective at reducing false positives detected by Mask R-CNN. It was also found that view selection along different axes of the 3D volume for model training and testing significantly impacted the Mask R-CNN performance with the best strategy being training a model on images from all XYZ views and making predictions only on images from the Z view. In addition, the amount of training data could be reduced to as low as 1 % without drastically reducing segmentation accuracy for our dataset. The proposed data-driven method has promising potential to facilitate RSA analysis in X-ray CT scans without the burden of large GPU requirements or manually annotating large datasets for model training.

## Data availability

Data and models are accessible at <https://auburn.box.com/s/l30xvb3h3mo5omjeyow35mg2i2tqjnhq>.

Source code can be found at <https://github.com/marybethcassity/3DRootSegmentation>.

## CRediT authorship contribution statement

**Mary E. Cassity:** Software, Methodology, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Paul C. Bartley:** Data curation, Funding acquisition, Supervision, Writing – review & editing. **Yin Bao:** Conceptualization, Methodology, Supervision, Investigation, Funding acquisition, Writing – review & editing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yin Bao reports financial support was provided by Alabama Agricultural Experiment Station. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This project was supported by the Alabama Agricultural Experiment Station. The authors would like to thank Spencer Overton and Austin Lindquist for supporting this project by performing manual annotation of the datasets.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.atech.2024.100666](https://doi.org/10.1016/j.atech.2024.100666).

## References

- [1] Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools.
- [2] Brooks, J. (2019). COCO Annotator. <https://github.com/jsbroks/coco-annotator>.
- [3] H.F. Downie, M.O. Adu, S. Schmidt, W. Otten, L.X. Dupuy, P.J. White, T. A. Valentine, Challenges and opportunities for quantifying roots and rhizosphere interactions through imaging and image analysis, *Plant Cell Environ.* 38 (7) (2015) 1213–1232, <https://doi.org/10.1111/pce.12448>.
- [4] R.J. Flavel, C.N. Guppy, S.M.R. Rabbi, I.M. Young, An image processing and analysis tool for identifying and analysing complex plant root systems in 3D soil using non-destructive analysis: root1, *PLoS One* 12 (5) (2017) e0176433, <https://doi.org/10.1371/journal.pone.0176433>.
- [5] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering, in: *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998. Proceedings 1*, Springer, Berlin Heidelberg, 1998, pp. 130–137, <https://doi.org/10.1007/BFb0056195>.
- [6] N. Gaggion, F. Ariel, V. Daric, É. Lambert, S. Legendre, T. Roulé, A. Camoirano, D. H. Milone, M. Crespi, T. Blein, E. Ferrante, ChronoRoot: high-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture, *GigaScience* (2021), <https://doi.org/10.1093/gigascience/giab052>.
- [7] T. Galkovskyi, Y. Mileyko, A. Bucksch, B. Moore, O. Symonova, C.A. Price, C. N. Topp, A.S. Iyer-Pascuzzi, P.R. Zurek, S. Fang, J. Harer, P.N. Benfey, J.S. Weitz, GIA Roots: software for the high throughput analysis of plant root system architecture, *BMC Plant Biol.* 12 (1) (2012) 116, <https://doi.org/10.1186/1471-2299-12-116>.
- [8] W. Gao, S. Schlüter, S.R.G.A. Blaser, J. Shen, D. Vetterlein, A shape-based method for automatic and rapid segmentation of roots in soil from X-ray computed tomography images: routine, *Plant Soil* 441 (1–2) (2019) 643–655, <https://doi.org/10.1007/s11104-019-04053-6>.
- [9] S. Gerth, J. Claußen, A. Eggert, N. Wörlein, M. Waininger, T. Wittenberg, N. Uhlmann, Semiautomated 3D root segmentation and evaluation based on X-ray CT imagery, *Plant Phenomics* (2021), <https://doi.org/10.34133/2021/8747930>.
- [10] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (2) (2020) 386–397, <https://doi.org/10.1109/TPAMI.2018.2844175>.
- [11] H. Janzen, Correcting mathematically for soil adhering to root samples, *Soil Biol. Biochem.* 34 (12) (2002) 1965–1968, [https://doi.org/10.1016/S0038-0717\(02\)00206-7](https://doi.org/10.1016/S0038-0717(02)00206-7).
- [12] J. Kang, L. Liu, F. Zhang, C. Shen, N. Wang, L. Shao, Semantic segmentation model of cotton roots in-situ image based on attention mechanism, *Comput. Electron. Agric.* 189 (2021) 106370, <https://doi.org/10.1016/j.compag.2021.106370>.
- [13] I.T. Koekoets, J.H. Venema, J.T.M. Elzenga, C. Testerink, Roots withstanding their environment: exploiting root system architecture responses to abiotic stress to improve crop tolerance, *Front. Plant Sci.* 07 (2016), <https://doi.org/10.3389/fpls.2016.01335>.
- [14] S.J. Livesley, Sieve size effects on root length and biomass measurements of maize (*Zea mays*) and *Grevillea robusta*, *Plant Soil* 207 (2) (1998) 183–193, <https://doi.org/10.1023/A:1026461107110>.
- [15] S. Mairhofer, S. Zappala, S.R. Tracy, C. Sturrock, M. Bennett, S.J. Mooney, T. Pridmore, RootTrak: automated recovery of three-dimensional plant root architecture in soil from X-ray microcomputed tomography images using visual tracking, *Plant Physiol.* 158 (2) (2012) 561–569, <https://doi.org/10.1104/pp.111.186221>.
- [16] S.J. Mooney, T.P. Pridmore, J. Hellawell, M.J. Bennett, Developing X-ray Computed Tomography to non-invasively image 3-D root systems architecture in soil, *Plant Soil* 352 (1–2) (2012) 1–22, <https://doi.org/10.1007/s11104-011-1039-9>.
- [17] M. Phalempin, E. Lippold, D. Vetterlein, S. Schlüter, An improved method for the segmentation of roots from X-ray computed tomography 3D images: routine v.2, *Plant Methods* 17 (2021) 39, <https://doi.org/10.1186/s13007-021-00735-4>.
- [18] J. Sander, M. Ester, H.-P. Kriegel, X. Xu, Density-based clustering in spatial databases: the algorithm gdsScan and its applications, *Data Min. Knowl. Discov* 2 (2) (1998) 169–194, <https://doi.org/10.1023/A:1009745219419>.
- [19] V. Satopaa, J. Albrecht, D. Irwin, B. Raghavan, Finding a "kneedle" in a haystack: detecting knee points in system behavior, in: *2011 31st International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171, <https://doi.org/10.1109/ICDCSW.2011>.

- [20] E. Schubert, J. Sander, M. Ester, H.P. Kriegel, X. Xu, DBSCAN revisited, revisited: why and how you should (still) use DBSCAN, *ACM Transactions on Database Systems* 42 (3) (2017) 1–21, <https://doi.org/10.1145/3068335>.
- [21] T. Selzner, J. Horn, M. Landl, A. Pohlmeier, D. Helmrich, K. Huber, A. Schnepf, 3D U-Net segmentation improves root system reconstruction from 3D MRI images in automated and manual virtual reality work flows, *Plant Phenomics* 5 (2023) 0076, <https://doi.org/10.34133/plantphenomics.0076>.
- [22] J.A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, 2nd ed., Cambridge University Press, 1999.
- [23] M.R. Shao, N. Jiang, M. Li, A. Howard, K. Lehner, J.L. Mullen, S.L. Gunn, J. K. McKay, C.N. Topp, Complementary phenotyping of maize root system architecture by root pulling force and X-ray imaging, *Plant Phenomics* 2021 (2021), <https://doi.org/10.34133/2021/9859254>.
- [24] C. Shen, L. Liu, L. Zhu, J. Kang, N. Wang, L. Shao, High-throughput in situ root image segmentation based on the improved DeepLabv3+ method, *Front. Plant Sci.* 11 (2020), <https://doi.org/10.3389/fpls.2020.576791>.
- [25] S. Smith, I. De Smet, Root system architecture: insights from *Arabidopsis* and cereal crops, *Philosophical Transactions of the Royal Society B: Biolog. Sci.* 367 (1595) (2012) 1441–1452, <https://doi.org/10.1098/rstb.2011.0234>.
- [26] M. Soltaninejad, C.J. Sturrock, M. Griffiths, T.P. Pridmore, M.P. Pound, Three dimensional root CT segmentation using multi-resolution encoder-decoder networks, *IEEE Trans. Image Process.* 29 (2020) 6667–6679, <https://doi.org/10.1109/TIP.2020.2992893>.
- [27] A. Tabb, K.E. Duncan, C.N. Topp, Segmenting root systems in X-ray computed tomography images using level sets, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 586–595, <https://doi.org/10.1109/WACV.2018.00070>.
- [28] V. Thesma, J. Mohammadpour Velni, Plant root phenotyping using deep conditional GANs and binary semantic segmentation, *Sensors* 23 (1) (2022) 309, <https://doi.org/10.3390/s23010309>.
- [29] E.W. Tollner, X-ray computed tomography applications in soil ecology studies, *Agric. Ecosyst. Environ.* 34 (1–4) (1991) 251–260, [https://doi.org/10.1016/0167-8809\(91\)90112-B](https://doi.org/10.1016/0167-8809(91)90112-B).
- [30] M. van Noordwijk, J. Floris, Loss of dry weight during washing and storage of root samples, *Plant Soil* 53 (1–2) (1979) 239–243, <https://doi.org/10.1007/BF02181896>.
- [31] van Rossum, G. (2022). *Python 3 Reference Manual*. <https://docs.python.org/releases/3.10.9/>.
- [32] T. Wang, M. Rostamza, Z. Song, L. Wang, G. McNickle, A.S. Iyer-Pascuzzi, Z. Qiu, J. Jin, SegRoot: a high throughput segmentation method for root image analysis, *Comput. Electron. Agric.* 162 (2019) 845–854, <https://doi.org/10.1016/j.compag.2019.05.017>.
- [33] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. <https://github.com/facebookresearch/detectron2>.
- [34] Q. Yu, H. Tang, L. Zhu, W. Zhang, L. Liu, N. Wang, A method of cotton root segmentation based on edge devices, *Front. Plant Sci.* 14 (2023), <https://doi.org/10.3389/fpls.2023.1122833>.
- [35] Zhou, Q.-Y., Park, J., & Koltun, V. (2018). Open3D: a Modern Library for 3D Data Processing. <http://www.open3d.org/>.