Master's Thesis Opportunity

Bioinformatics • University of Potsdam / Helmholtz-Zentrum Berlin

Automated 3D Segmentation of Arabidopsis thaliana Roots and Root Hairs from High-Resolution X-ray CT Data

Degree: Master of Science (Bioinformatics or related field)

Location: University of Potsdam / Helmholtz-Zentrum Berlin, Germany

Supervisors: Dr. Jacqueline Nowak (University of Potsdam) and Boyana Kozhuharova (University of Potsdam)

Contact: janowak@uni-potsdam.de • boyana.kozhuharova.1@uni-potsdam.de

Background

Root hairs are slender epidermal outgrowths of Arabidopsis thaliana, typically 5–10 µm in diameter and up to 1 mm long. Despite representing only ~2% of root biomass, they dramatically expand the root absorptive surface and play critical roles in phosphorus and iron uptake, drought tolerance, and soil anchorage. Understanding their three-dimensional organisation in natural soil conditions is therefore of high biological and agricultural relevance.

High-resolution X-ray computed tomography (CT) enables non-destructive, in situ 3D imaging of roots and root hairs growing in sand or soil. However, automated segmentation of these fine, low-contrast structures from volumetric CT data remains a major methodological challenge. An automated Python-based pipeline has been developed and benchmarked against VGSTUDIO MAX across six CT datasets from three Arabidopsis genotypes with distinct root hair densities (Sav-0, Mt-0, Col-0), achieving a mean Dice Similarity Coefficient (DSC) of 0.78 for main root segmentation. While main root segmentation is robust under favourable imaging conditions, consistent root hair detection remains an open problem — limited by low contrast, residual water artefacts, and the current slice-wise tracking strategy.

This Master's thesis project will directly continue and extend this work, combining improvements to sample preparation, imaging protocols, and the segmentation pipeline itself, with the goal of achieving reliable, automated 3D root hair segmentation. The project also offers a unique opportunity to contribute to an emerging high-throughput phenotyping effort: the X-ray Tomography Lab at Helmholtz-Zentrum Berlin is actively working to expand the automation of its imaging workflows through the integration of robotics and large language models (LLMs). Depending on project progress, the student may have the chance to contribute to the establishment of the first protocols for high-throughput CT-based root hair phenotyping.

Project Aims

The thesis comprises four interconnected workstreams that together address both the experimental and computational bottlenecks of the current approach:

1. New Data Generation

A key component of the thesis will be the acquisition of new CT datasets under improved experimental conditions. This includes:

- Imaging additional biological replicates of the existing genotypes (Sav-0, Mt-0, Col-0) with optimised sample preparation to expand the benchmark dataset
- Potentially extending the genotype panel to new Arabidopsis accessions with distinct root hair morphologies, broadening the biological scope of the project and challenging the pipeline's generalisability across a wider range of phenotypes
- Generating ground-truth segmentations for new datasets to support both pipeline evaluation and future deep learning model training

2. Improved Sample Preparation and Contrast Enhancement

The primary limitation identified in the existing pipeline is insufficient contrast between root hair tissue and the surrounding background. The student will test and evaluate contrast-enhancing staining agents, with priority given to:

- Phosphotungstic acid (PTA) — tungsten-based, strong and homogeneous tissue penetration, predicted ~26% contrast gain at 33 keV
- Cesium iodide (CsI) — selectively binds to pectin-rich cell walls, offering specificity to root hair structures (~10% gain at 33 keV)

3. Algorithmic Improvements to the Python Segmentation Pipeline

Three specific failure modes have been identified that the student will address with targeted algorithmic improvements:

3a. Multiscale Hair Enhancement and Context-Aware Thresholding

To reduce dependence on ultra-conservative global intensity thresholds that fragment fine root hairs, the student will:

- Implement a multiscale Hessian filter (Frangi "vesselness" or Sato "tubeness") to enhance tubular, hair-like structures and suppress background
- Modify the root tracking algorithm to use the validated root from the previous slice as a template, extracting only the portion of merged components that aligns with it — preventing mergers from propagating into the background
- Exploit the improved tracking to relax intensity thresholds, recovering faint root hair voxels without introducing large-scale false positives

3b. Post-Tracking Root Hair Fragment Linkage

The current tracker discards root hair fragments that are not continuously connected to the main root across slices. To recover these systematically, the student will:

- Harvest small, hair-scale connected components within a defined distance band around the segmented root
- Compute a multiscale hair-likeness map (using vesselness/tubeness filters) to score candidate voxels

- For each retained fragment, find the shortest path back to the main root within the distance band, preferring hair-like voxels, and retain connections that pass plausibility checks (path length, curvature, hair-likeness score)
- Skeletonize accepted connections, thicken to expected hair width, and add to the final segmentation mask

3c. Sand-Anchored Slice Normalisation for Artefact Correction

Slice-specific “dark veil” artefacts cause sudden grey-value shifts that prevent consistent thresholding. The student will:

- Implement a sand-anchored affine normalisation: select a veil-free reference slice, record robust sand intensity landmarks (median, 10th/90th percentiles), and rescale each affected slice so that sand intensities match the reference
- Apply light smoothing along the Z-axis to prevent over-correction, and test against the previously trialled (and problematic) full histogram matching approach
- Evaluate whether a coarse background subtraction (rolling-ball) can remove residual low-frequency shading without affecting fine structural detail

4. Deep Learning Segmentation (Exploratory)

As a complementary and forward-looking component, the student will explore whether a small deep learning model can outperform the rule-based pipeline on this dataset:

- Train a 3D U-Net using patch-based sampling on the available CT volumes, using VGSTUDIO MAX segmentations as pseudo-labels and a hand-corrected subset for fine-tuning
- Apply strong data augmentation (intensity shifts, blur, noise) and cross-validation by stack to maximise generalisation from a small dataset
- Evaluate performance on held-out stacks and compare against the improved classical pipeline
- Assess whether the classical pipeline’s outputs can serve as a scalable annotation tool to generate future training data

Broader Context: Towards High-Throughput Phenotyping

This project sits at the interface of plant biology and emerging automation technologies. The X-ray Tomography Lab at Helmholtz-Zentrum Berlin is actively pursuing the integration of robotic sample handling and large language models (LLMs) into its imaging workflows, with the long-term goal of enabling high-throughput, largely unsupervised CT-based phenotyping. Depending on the progress of the thesis work, there may be an exciting opportunity for the student to contribute to the establishment of the first protocols for high-throughput root hair phenotyping within this framework — a contribution that would position this work at the forefront of next-generation plant phenomics.

Expected Deliverables

- New CT datasets acquired under optimised imaging conditions, potentially including new Arabidopsis accessions
- Optimised sample preparation protocol for high-contrast CT imaging of Arabidopsis root hairs
- Extended and improved Python segmentation pipeline, incorporating multiscale filtering, topology-aware tracking, fragment linkage, and artefact-robust normalisation

- Quantitative evaluation of pipeline improvements against the existing and newly acquired VGSTUDIO MAX ground truth
- Exploratory 3D U-Net segmentation model with performance comparison
- Master's thesis and, where results warrant, contribution to a follow-up scientific publication

Candidate Profile

Required

- Enrolled in or about to complete a Master's degree in Bioinformatics, Computational Biology, Computer Science, or a closely related field
- Solid programming experience in Python, including familiarity with NumPy, SciPy, and scikit-image
- Understanding of image processing fundamentals (filtering, thresholding, morphological operations, connected components)
- Ability to work independently and document code and methods clearly

Beneficial (not required)

- Experience with 3D image analysis or volumetric data (e.g. CT, MRI, microscopy stacks)
- Familiarity with deep learning frameworks (PyTorch or TensorFlow) and segmentation architectures (U-Net)
- Basic wet lab skills or willingness to learn standard plant biology protocols (seed sterilisation, seedling cultivation)
- Knowledge of X-ray CT principles or other tomographic imaging methods
- Experience with scikit-learn, PyTorch, or medical/biological image analysis libraries (e.g. SimpleITK, napari)
- Familiarity with version control (Git/GitHub)

Personal qualities

- Curiosity about the intersection of computational methods and plant biology
- Methodical approach to debugging and benchmarking
- Comfortable working with large data volumes and long-running computations
- Interest in contributing to open, reproducible science and emerging automation technologies

What We Offer

- A well-defined project with existing data, code, and ground-truth annotations — you will not start from scratch
- Access to high-resolution X-ray CT infrastructure at Helmholtz-Zentrum Berlin (HZB)
- Interdisciplinary supervision spanning bioinformatics, plant biology, and X-ray imaging
- Exposure to cutting-edge automation efforts integrating robotics and LLMs into CT imaging workflows
- A collaborative, international research environment across two leading research institutions

- Flexible start date and possibility to shape the project based on your own strengths and interests
- Opportunity to be included as an author on a publication that may result from this work

How to Apply

Please send the following to janowak@uni-potsdam.de and boyana.kozhuharova.1@uni-potsdam.de:

- A short motivation letter (max. 1 page) explaining your interest in the project and relevant experience
- Your CV
- Optional: a link to a GitHub repository, code sample, or previous project report

Applications are reviewed on a rolling basis. We encourage early applications as the position will be filled as soon as a suitable candidate is found.

We look forward to hearing from you.