

SegPlantFormer: A Transformer-Based Model for Plant Semantic Segmentation

Cristóbal Alcázar¹*, Edward Chocano²*, and Ricardo Flores²*, *equal contribution

¹Universidad de Chile, CL; ²Worcester Polytechnic Institute, US

Correspondence to: cristobal.alcazar@ug.uchile.cl, ejchocanocoralla@wpi.edu, rflores@wpi.edu

github.com/alcazar90 — @vamos_alcazar



In a nutshell

- **Relevance.** Microscopic imaging has been pivotal in understanding biological and medical phenomena, providing insights into cellular structures, tissues, and organisms.
- **Current challenges.** Lack of labelled datasets for microscopic image analysis in plant cells.
- **Proposal.** A Transformer-based method for image segmentation to automate screening (detection of live/dead plant cell instances). Train on a dataset created by our team and carefully annotated by human experts.
- **Validation.** Comparing the outcomes with the traditional method for screening plant cells based on image processing using colour saturation techniques, as employed in the state-of-the-art studies in this field.

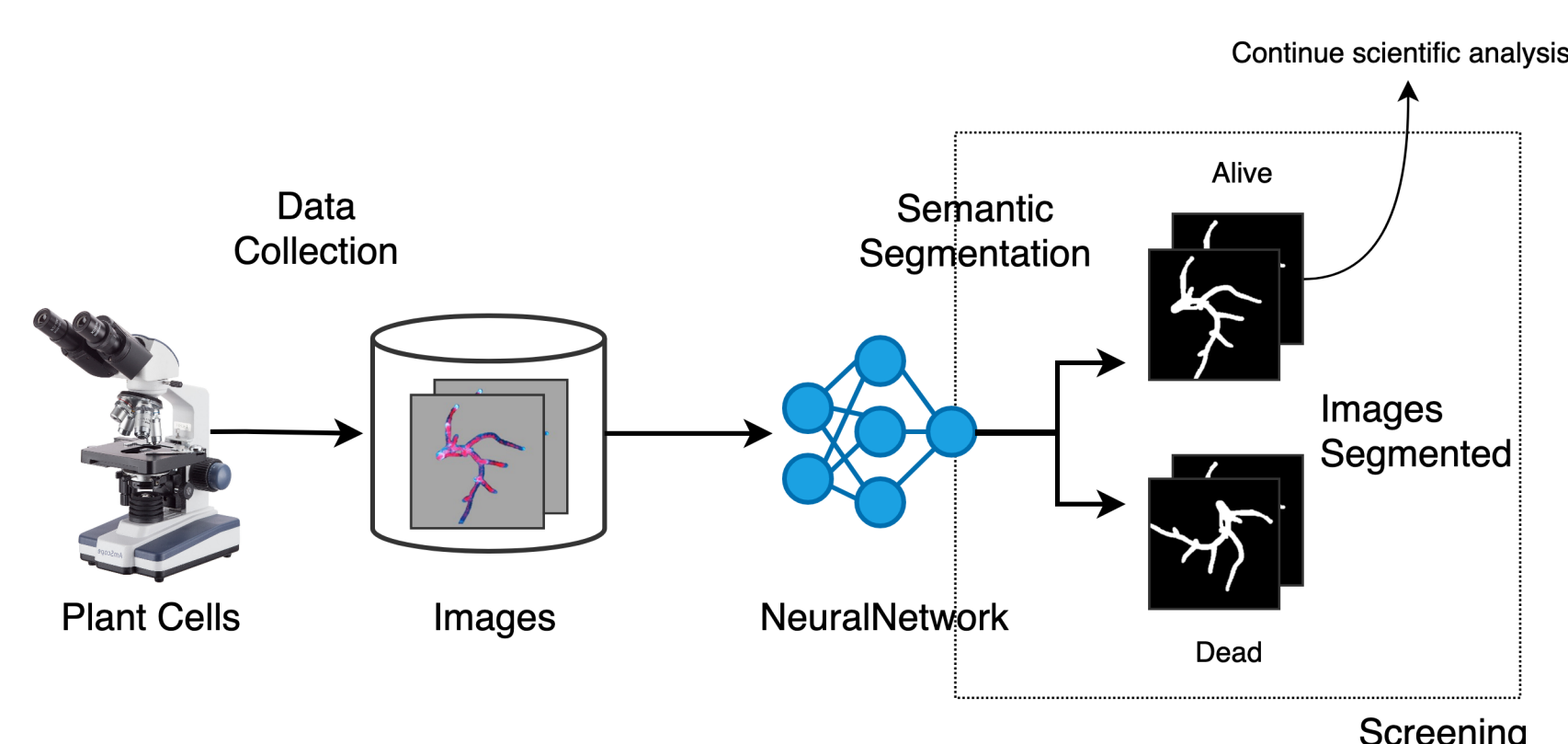


Fig. 1. Project pipeline.

Dataset Collection

- The dataset spans eight days, targeting *P. patens* moss growth patterns, merging images from three experiments into stacked images.
- Initial dataset phases include "Control WildType (CWT)" and "dead" images, representing living and non-living cells, respectively.
- Stacked images are broken down into smaller patches (A) and (B), represented as white squares.
- The input images (A1) and (B1) are denoted as \mathcal{X} in the figure.

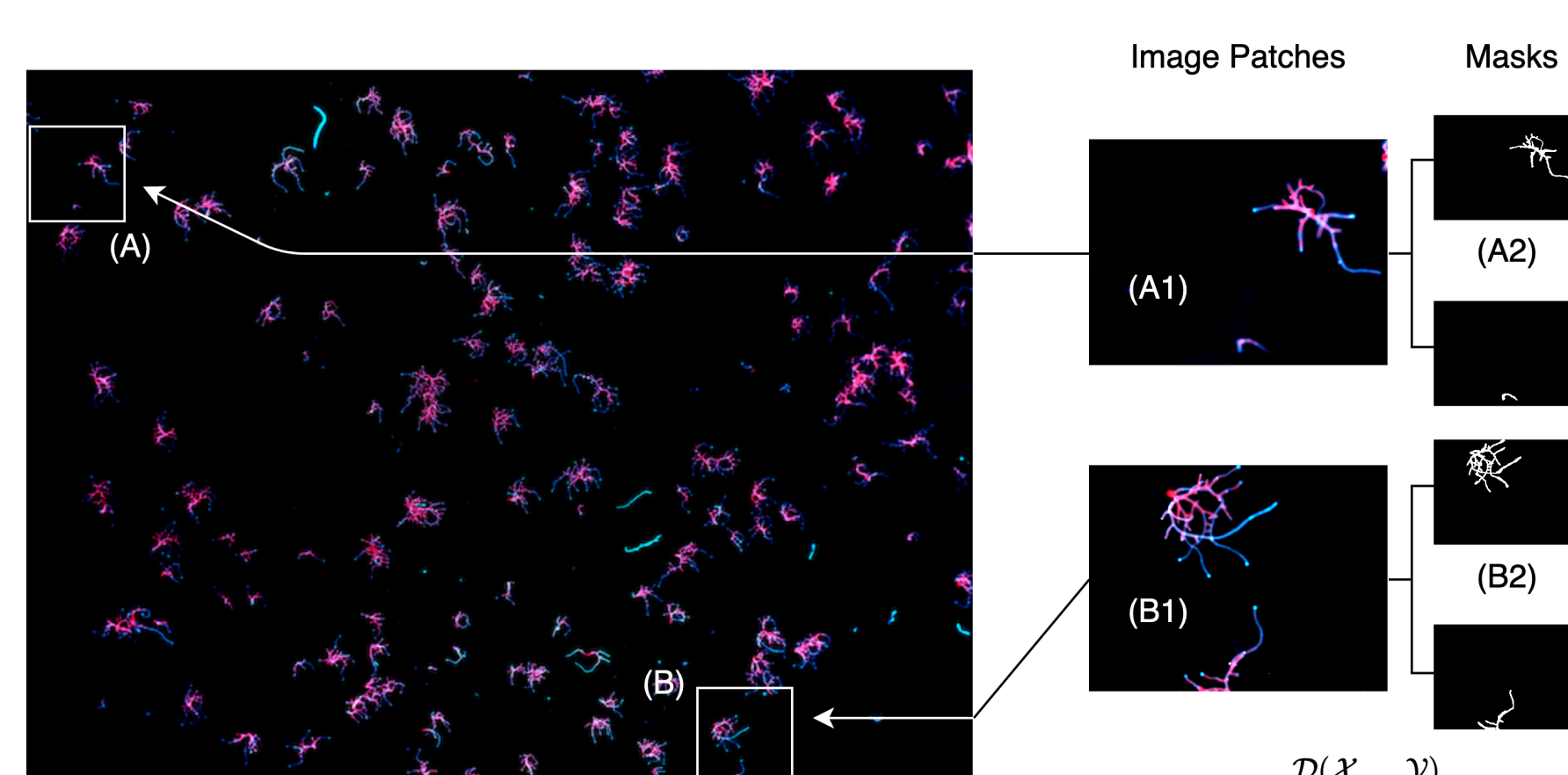


Fig. 2. **Left:** The results of the laboratory experiments are captured in a large stacked image (25036×18954 px) with high-resolution microscopic imaging. The white squares with labels A and B are samples that were taken. **Right:** Smaller image patches from the samples, such as (A1) and (B1), are the input images for the model. Then, we manually handcraft masks with semantic information from these images regarding the plant shapes associated with alive/dead cell labels (A2 and B2), elaborating the dataset \mathcal{D} .

Data labelling with Label Studio

We used Label Studio to create segmentation masks (\mathcal{Y}) guided by expert knowledge for each image. This manual annotation process produced one or more masks for each image patch (A2) and (B2).



Fig. 3. An example of labelling an CWT image with the **Label Studio** Program. In addition to labelling living and dead plants, we generated tags to identify instrument-generated noise and plants cut during the image patch extraction process. Although not initially used, this information holds value in enhancing the system's effectiveness.

A Transformer-Based model for Image Segmentation

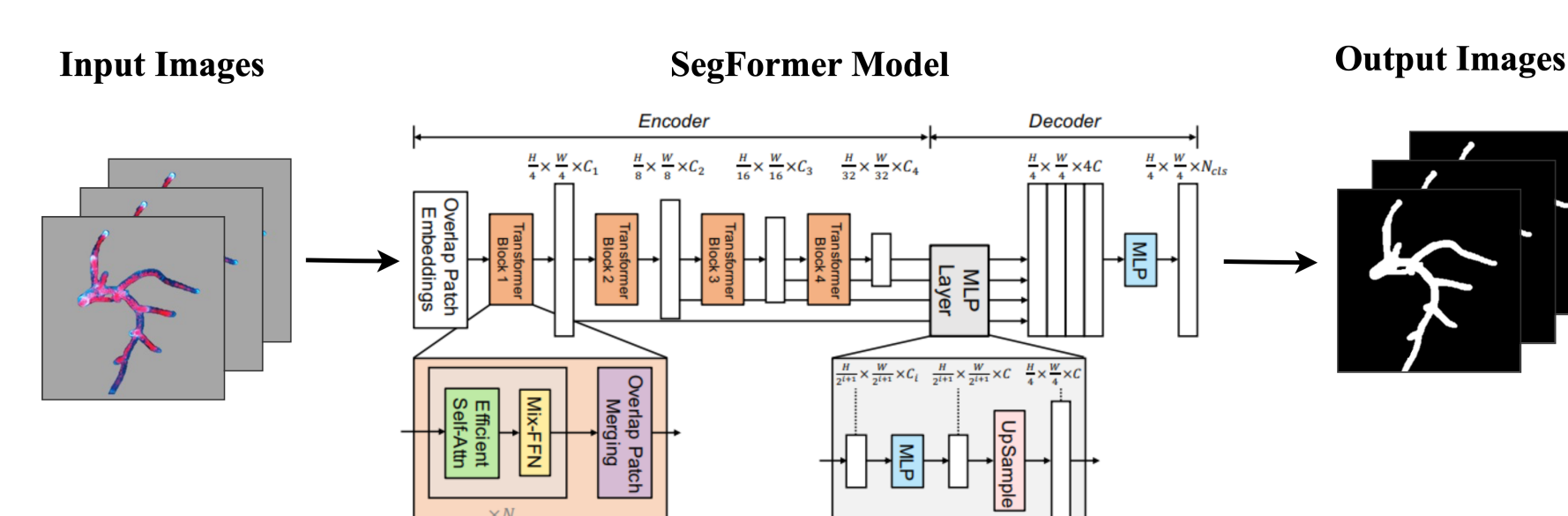


Fig. 4. **SegPlantFormer Framework** Model architecture. An illustration of a forward pass of a cell plant image to obtain a segmentation mask. Source (Xie, Neurips-2021).

- We finetuned two models, one for the **CWT** and the other for the **dead** datasets, to perform binary segmentation to detect the class or not.

Results

- Our method demonstrates a 27.08% superior performance over the conventional method for live plants.
- The experiments exhibit consistent results, displaying a standard deviation of 0.005 and an average mIoU of 0.535 for the **CWT** dataset.
- Our method achieves a 67.19% improvement in mIoU for the **dead** dataset compared to the traditional approach.
- The **dead** dataset is characterized by higher complexity and heightened noise levels than the CWT.

Dataset	Methods		
	Traditional	Our	Gain (%)
CWT	0.421	0.535 ± 0.005	27.08
dead	0.256	0.428 ± 0.029	67.19
Avg.	0.338	0.481	0.143

Table 1. **Models Result:** average \pm standard deviation of Mean Intersection Over Union (*mIoU*) scores from fine-tuning SegFormer on cwt and dead datasets five times.

Conclusions

- **SegPlantFormer** surpasses traditional plant cell screening methods to detect alive/dead plants.
- **It utilizes data's power for adaptability**, enabling continuous improvement and addressing challenges like overlapping, currently impossible with the traditional method.

Future Work

- Expand dataset with new experiments and tasks like phenotype labels and dense estimation.
- Share the dataset with the scientific community to encourage collaboration.
- Address the challenge of overlapping in dense estimation tasks.
- Develop a comprehensive toolkit for streamlined and automated screening.

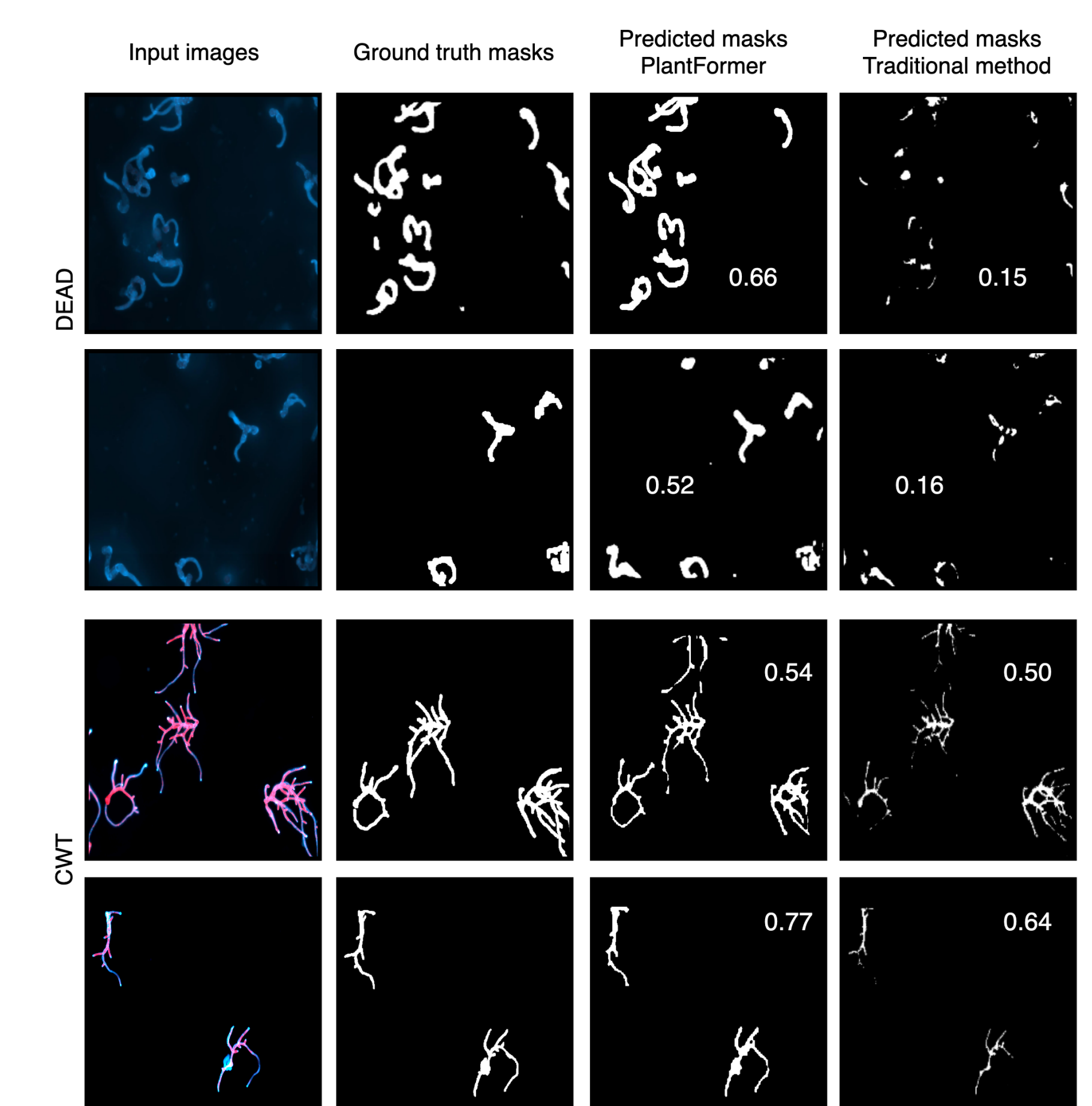


Fig. 5. Compares approaches between two subsets: the top rows display input images from the **dead** dataset, and the bottom rows depict from the **CWT** dataset. The IoU metrics are displayed within the images in columns 3 and 4. The model consistently outperforms the traditional method in almost all test cases, as seen in the first row, where the predicted mask achieves 0.66 IoU compared to the traditional method's 0.15 IoU. These observations were not used to train the model.

References



QR code for paper references