Machine Learning Aided Inner Ear Segmentation

Katya Luchette, Advisor: Jan Mulder

Abstract

Segmentation is the process of identifying regions of interest in images. This process is commonly utilized with medical images to identify regions such as tissues, tumors, or organs for more effective diagnoses or personalized analyses. Segmentation can also be a useful tool for producing 3D visualizations of various organs. Current segmentation methods primarily rely on manual segmentation processes, which can be time-consuming and labor-intensive. This project presents a workflow for automating the segmentation process by using machine learning algorithms to segment out regions in tissue images of the inner ear.

1 Introduction

Medical image segmentation has many practical applications to the medical field, including disease diagnosis and treatment planning[1]. Manual segmentation is a primary method for extracting regions of interest from medical images. The process involves a user manually outlining structures of interest in slices of radiological tissue data[2]. This method is often limited by its high time consumption and dependence on expert knowledge[2]. Additionally, the manual segmentation process allows for limited reproducibility and is imprecise due to user bias[2].

Machine learning presents a potential solution to these challenges in its ability to streamline the process and produce segmented image results at a much faster rate than the manual process. This project describes a workflow for using machine learning algorithms to identify two separate regions in the inner ear: the nucleus and the cell body. The model segmented images are then compressed to form an interactive 3D visualization of the inner ear.

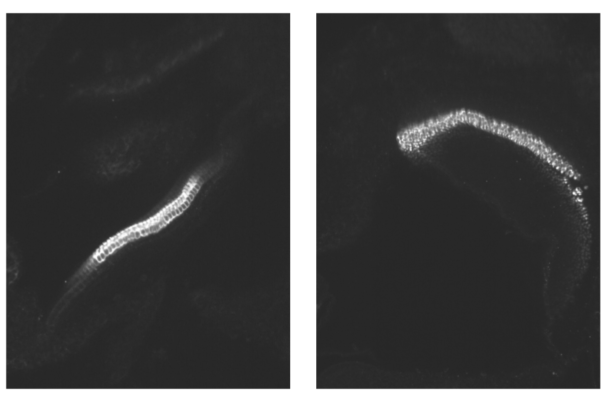

Figure 1: Example tissue images from the inner ear.

2 Methods

2.1 Data Collection

The dataset used for training was provided by Karolinska Institute. The dataset contained 1183 grayscale tissue slice images of the inner ear. To train the model, 44 images were chosen from sections across the 1183 image range. For the initial rounds of training, 10 images were chosen from a single region of the inner ear. This resulted in a model that was not able to generalize well to differing regions of the inner ear. To increase variability in the ground truth dataset, a few images were randomly chosen approximately every 100 images in the data stack, resulting in the final 44 image ground truth dataset. This allowed for more variability in the ground truth set and ensured that the model was able to generalize well across all images in the inner ear stack.

For each grayscale image, we produced a corresponding manual segmentation for the separate regions of interest using Napari. Thus, there were an additional 44 mask images which identified regions in the grayscale image which corresponded to nuclei and an additional 44 mask images which identified regions which corresponded to cell bodies.

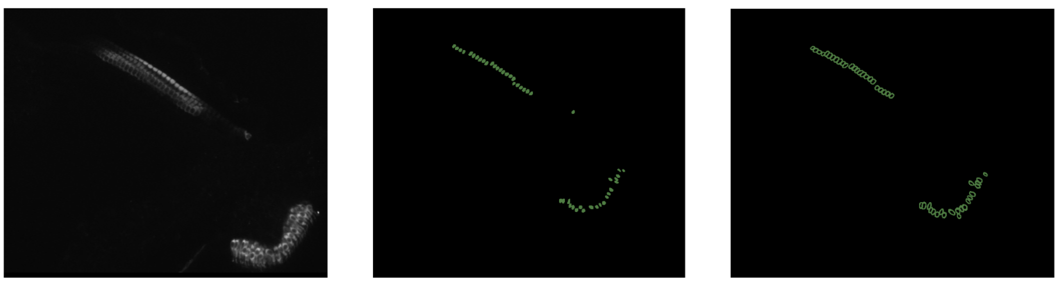

Figure 2: A ground truth inner image and its corresponding nucleus manual segmentation mask (middle) and cell body manual segmentation mask (right).

2.2 Data Preprocessing

After the manual segmentation process was completed, each mask was converted into a binary mask. This converted the mask image into a computer-interpretable bitwise mask, where regions of interest were shown in white and background information was shown in black. The grayscale images and their corresponding masks were then loaded into an image stack and label stack, respectively. This ensured that each image was correctly paired with its corresponding binary mask and could be processed efficiently. All images and labels were then resized to 2048 x 2048 pixels to ensure equal proportions. To artificially increase the size of the small dataset, a process called patchify was used during which each image was further divided up into smaller sections. For both the nucleus and cell body segmentation process, it was found that a patchify size of 512 allowed the model to identify nuclei and a size of 256 allowed the model to identify cell bodies successfully. Each image and mask was then normalized to ensure consistent pixel values across all images.

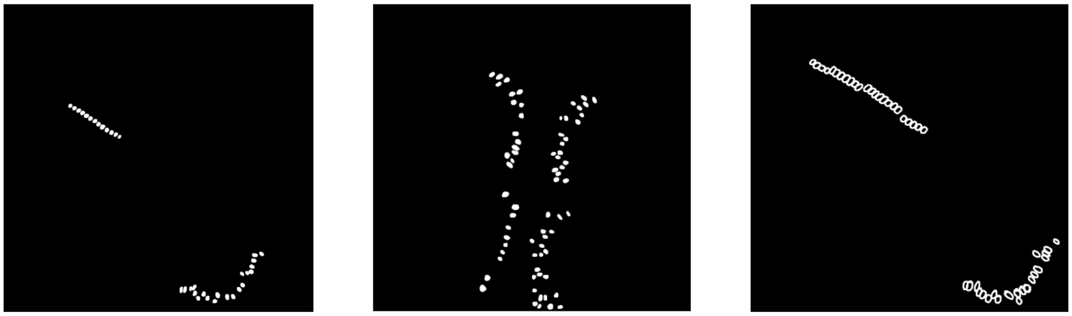

Figure 3: Example binary masks.

2.3 Machine Learning Model

The data preprocessing and machine learning model are based on a modified version of S. Bhattiprolu's code for semantic segmentation with a small dataset[3]. A UNet model is a neural network commonly used for image segmentation. It classifies each pixel in an image as belonging to a region of interest or as belonging to the background. Because of the small size of the dataset and UNet's capabilities for producing semantic segmentation, a pretrained UNet model was used for training. Two separate models were trained: one for nucleus segmentation and one for cell body segmentation.

3 Results

The machine learning models were able to successfully identify nuclei and cell bodies when given images of the inner ear. While the nucleus model created smoother segmentations of nuclei, the cell body model was still able to identify cell bodies with some level of noise. The cell body model achieved an IoU score of 0.44, while the nucleus model achieved an IoU score of 0.55. An IoU score measures the accuracy of overlap between a ground truth image and a predicted image on a scale from 0 to 1. These scores reflect a moderate level of accuracy in the overlap between the ground truth and prediction images, thus there is room for improvement in future iterations of these models.

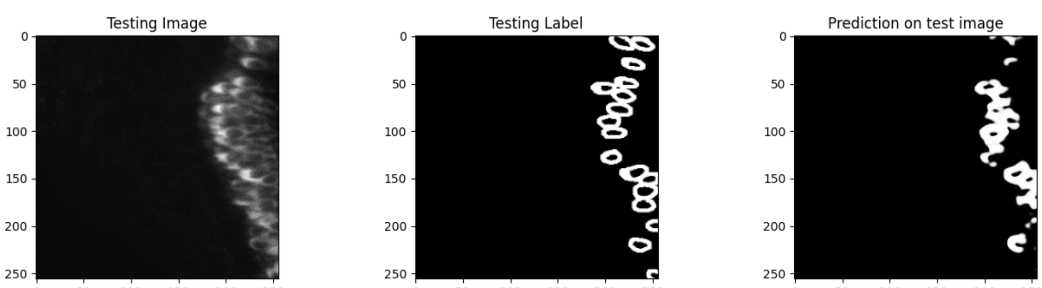

Figure 4: Comparison between manually segmented cell bodies (middle) and model segmented cell bodies (right).

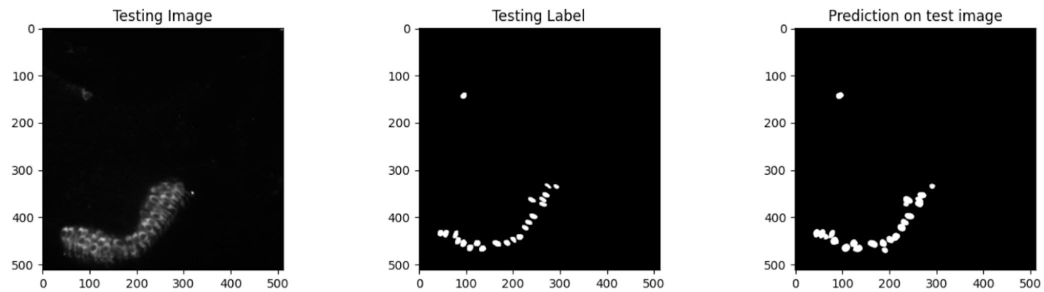

Figure 5: Comparison between manually segmented nuclei (middle) and model segmented nuclei (right).

4 Discussion

The machine learning model was able to run 1183 predictions in approximately 20 minutes, which is a much faster rate than the manual segmentation process. This demonstrates that machine learning-aided segmentation is a viable and efficient alternative to manual methods. Furthermore, the segmented images generated by the model contained the necessary information to produce 3D reconstructions of the inner ear, making these models valuable tools for both analysis and visualization.

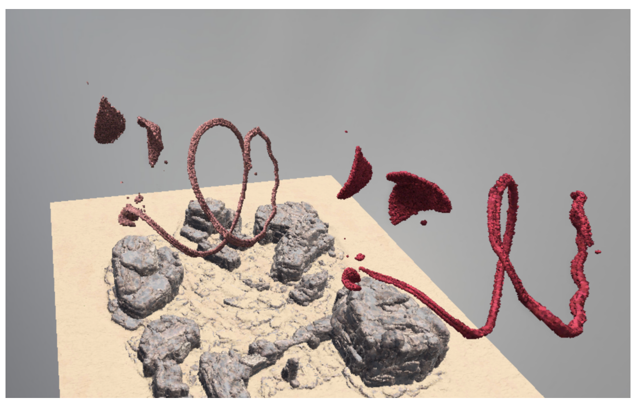

Figure 6: 3D visualization of model segmentations of the inner ear.

To further improve this project, exploring alternative models could increase the accuracy of the predictions. Additionally, challenges were encountered with producing accurate Z-level dimensions during the visualization process, and thus more precise Z-level calculations could be used to address this issue. Lastly, a more robust approach to smoothing and upscaling images could be introduced to reduce noise in predictions, leading to better quality segmentations.

5 Acknowledgements

I would especially like to thank Jan Mulder for his support and mentorship during this project. I am also thankful to Evelina Husén for her guidance and encouragement. Furthermore, I extend my appreciation to the Karolinska Institute for providing the resources that made this project possible, as well as Atomontage for transforming this project into an interactive visualization.

6 GitHub

`https://github.com/luchetk/innerearsegmentation.git`

References

[1] 3D and Quantitative Imaging Laboratory, Technique of the Week – Segmentation, *3D and Quantitative Imaging Laboratory*, 2024, `https://3dqlab.stanford.edu/technique-of-the-week-segmentation/`.

[2] Bernhard Preim and Charl Botha, Chapter 4 - Image Analysis for Medical Visualization, in *Visual Computing for Medicine*, 2014, 111-175, `https://www.sciencedirect.com/science/article/abs/pii/B9780124158733000043`.

[3] Sreenivas Bhattiprolu, Mito UNet Transfer Learning with 12 Training Images, *Python for Microscopists*, GitHub, accessed December 4, 2024, `https://github.com/bnsreenu/python_for_microscopists/blob/master/216_mito_unet__xferlearn_12_training_images.py`.