Draft version January 6, 2026
Typeset using LaTeX **twocolumn** style in AASTeX631

Lightweight 3D Reconstruction of Large Fluorescence Microscopy Volumes via Statistical and Deep Learning Segmentation

Thomas Lund[1] and Jan Mulder[2]

[1] *Carleton College*
[2] *Karolinska Institute*

(Dated: January 2026)

1. INTRODUCTION

Recent improvements in z stacked imaging methods such as confocal laser scanning microscopy and light sheet fluorescence microscopy have made large and high quality datasets available for many biological objects of interest. These datasets are crucial for various applications, including protein expression studies, blood vessel mapping, and nerve signal tracing. Because of the computational difficulty of rendering these datasets, their study is often restricted to those with access to high performance compute resources, in particular system memory, with typical rendering techniques often requiring in the hundreds of gigabytes to hold the full models efficiently. In this study, we focus on a three channel, volumetric 2056x2048x2022 dataset derived from a fluorescence microscopy scan of a mouse heart Uhlén et al. (2015), where blood vessel walls and nerve fibers were stained. We present a preprocessing and segmentation pipeline that converts multi-channel fluorescence images into voxelized binary masks corresponding to structures of interest. Similar works have also attempted to segment z stacked imaging on smaller datasets through varying segmentation methods. Zekri & Lang (2024) Our approach enables efficient downstream analysis and visualization of large-scale volumetric data on more modest computational hardware. This strategy facilitates access to whole organ imaging datasets and supports further analysis of complex biological structures.

2. INSTRUMENTAL LIMITATIONS

Multiplex immunohistochemistry and immunofluorescence (mIHC/IF) are powerful techniques for visualizing multiple protein markers within tissue sections, enabling the detailed study of complex biological structures. In this study mIHC/IF were employed to visualize protein markers in tissue sections, with each image representing the summed average of several frames. This approach precluded the deconvolution of the exact point spread function. Because each fluorescent channel was captured separately, the resulting frames did not align perfectly in the z-axis, further complicating the accuracy of spatial registration. Additionally, variations in focus due to the tissue's distance from the scanner caused a progressive loss of sharpness toward the periphery of each frame. A further complication arose from the heart wall, which trapped fluorescent chemicals during staining, leading to false-positive signals in the blood vessel scans. Finally, the intensity of each image varied, and there is an uneven illumination pattern within images, leading to inconsistent bright and dark portions of each slice. These factors necessitate significant pre-processing and require a more robust approach than an intensity threshold.

3. DATA PIPELINE

To reduce the data footprint while maintaining the full spatial accuracy, we first convert the multi-channel images into voxelized binary masks containing only voxels which had the features of interest.

3.1. *Preprocessing*

We begin by normalizing the intensity values across the z-direction of the full dataset one channel at a time to reduce image to image intensity variance. This normalization helps ensure consistency across slices of the z-stack and mitigate any depth-dependent variations in signal intensity. Next, we apply an inverse hyperbolic sine (asinh) filter to stretch each slice, which effectively preserves relative intensity levels before we compress the dynamic range[fig.3]. We then apply a Contrast Limited Adaptive Histogram Equalization Zuiderveld (1994) (CLAHE) to improve local contrast and flatten background intensities. CLAHE operates by enhancing the contrast within small local regions of the image, which scales each local intensity to a standard range across the whole image, evening out the slice, and enhancing contrast between background and fluorescing features[fig.2]. To maximize the intensity flattening, we set the clipping limit to 0.01 and used a window size of 32x32 pixels.

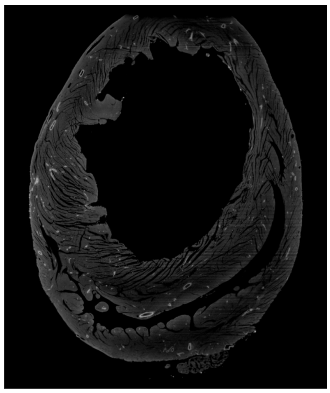

Figure 1. Scaled image slice before CLAHE

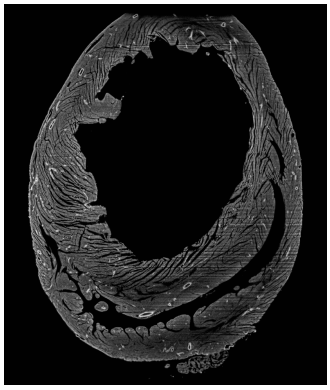

Figure 2. Scaled image slice after CLAHE

Finally, we use a robust scaler to remove outliers and normalize the data to the final 0-255 greyscale range.

3.2. *U-Net*

For image segmentation, we use a 9-layer U-Net architecture, which is well-suited for biomedical image segmentation tasks due to its ability to capture both local and global features Ronneberger et al. (2015). The model accepts input images of size 256x256 pixels. To prepare the original dataset, we tile the data slices into smaller 256x256 pixel tiles, which allows the model to handle high-resolution images efficiently. We manually segmented a small set of 32 tiles to create ground truth masks for training to test this model. This limited training set was used to evaluate the feasibility of the approach, a larger training dataset would likely increase performance. To supplement the size of the training set and improve model generalization, we applied standard data augmentation like rotation, translation and noise generation. For training, we used Tversky-Focal loss as our loss function. This loss function is designed to handle class imbalance, where the background often dominates the foreground structures, like is the case for potentially single-pixel nerve fibers or capillaries. To improve fitting performance we used a training loop with early stopping at loss plateauing, learning rate adaptation for oscillating loss, and model checkpoints that save the best model based on loss values on the test set. Training was carried out on 50 epochs, and the early stopping filter did not trigger. We had a best fit at loss = 0.55, and training accuracy = 0.98. Despite achieving high overall accuracy, small structures such as capillaries and nerve fibers were not segmented effectively by the model. To address this issue, we applied a difference of median filter, which helped to enhance the segmentation of these finer structures.

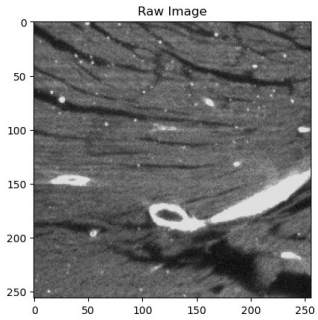

Figure 3. 256x256 input to U-Net model

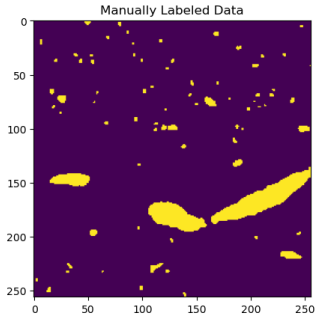

Figure 4. Manually labeled mask taken from the validation set

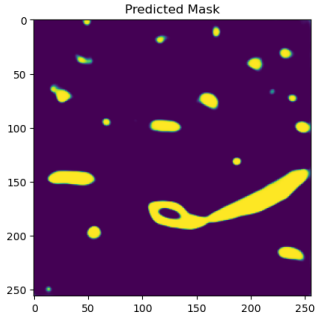

Figure 5. Predicted mask for the same image

3.3. *Difference of Medians*

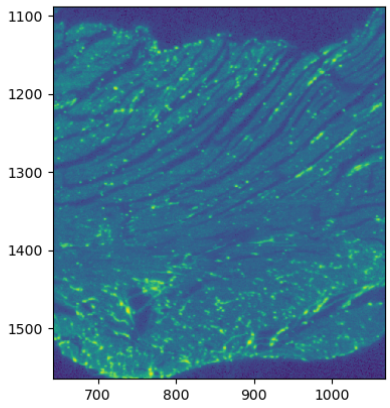

Figure 6. Zoomed section of a slice from the nerve fiber imaging

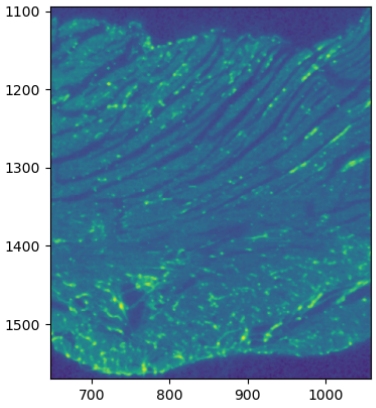

Figure 7. The small median image of the same slice (3x3)

To enhance medium-scale fluorescent structures in the grayscale images, we apply a difference-of-medians filtering approach designed to suppress both large-scale background features and high-frequency noise. For each image slice, we compute two median-filtered versions: one using a large kernel[fig.8] (15 × 15 pixels) and a second using a small kernel[fig.7] (3 × 3 pixels). The large-kernel median filter captures slowly varying, large-scale intensity components, such as organ boundaries and broad illumination gradients, while effectively averaging out smaller structures of interest. This filtered image therefore serves as an estimate of the local background intensity. Subtracting the large-kernel median image from the original slice yields an intermediate image in which small- and medium-scale fluorescent features are emphasized relative to their local background. However, this operation also amplifies single-pixel noise artifacts and hot pixels, which exhibit high local contrast and are not suppressed by the background subtraction alone. To mitigate this effect, we compute a second residual by subtracting the small-kernel (3 × 3) median-filtered image from the original slice, producing an estimate of high-frequency noise. Finally, subtracting this noise estimate from the background-subtracted image suppresses isolated hot pixels while preserving contiguous medium-scale fluorescent structures[fig.9]. The resulting image preferentially highlights features of interest, such as capillaries and nerve fibers, while reducing contributions from both large-scale background variations and pixel-level noise.

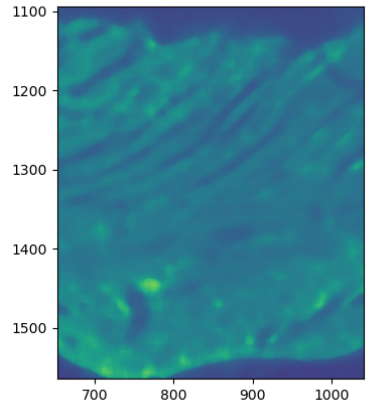

Figure 8. The large median image of the same slice (15x15)

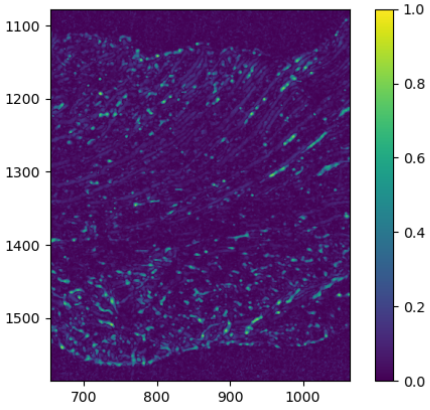

Figure 9. The subtracted I_3 and I_{15}

3.4. *Postprocessing*

Following preprocessing and segmentation, both pipelines yield a volumetric stack of binary images rep-

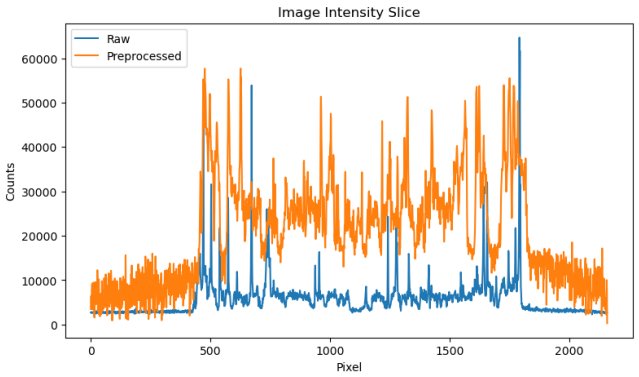

Figure 10. The raw and pre-processed intensities of a 1d slice of data through the x axis

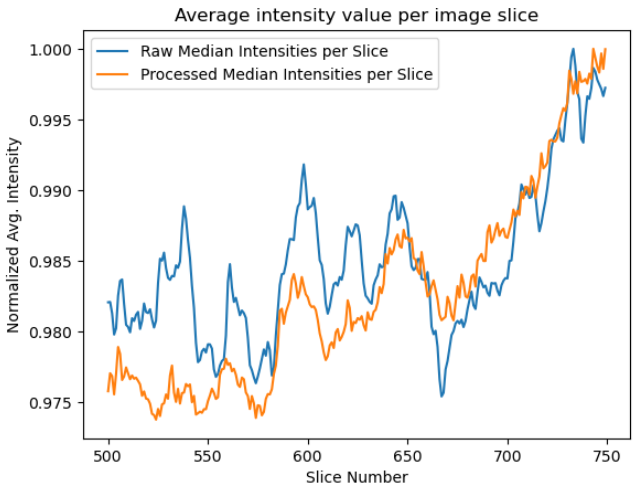

Figure 11. Average intensity profiles across a subset of slices before and after preprocessing. Preprocessing reduces inter-slice variability while preserving large-scale structural features.

resenting the structures of interest. Here, to take advantage of the well-defined spatial format and to clean up the data, we apply a kerneled erosion/dilation filter. First, to connect structures over the sparser z axis, we used a 1x1x5 kernel in dilation then erosion for 5 iterations. Then, to connect the shapes across the xy axis, we use a standard 3x3x3 kernel dilation/erosion loop once. To enable efficient visualization and downstream analysis, these binary volumes are converted into an optimized mesh representation with the marching cubes algorithm Lorensen & Cline (1987). The resulting meshes can be rendered efficiently on standard hardware and provide a compact representation of the original segmented volume. Finally, to reduce footprint even further and remove residual noise, we apply a number of faces filter, where any floating shape with fewer than 20 faces was removed.

4. RESULTS

4.1. *Preprocessing Results*

Applying intensity normalization and the asinh transform produced slices with consistent intensity profiles, reducing z-axis variation across the stack. Subsequent CLAHE effectively enhanced local contrast, making small fluorescent structures more distinct against background tissue.

Qualitatively, structures such as capillaries and nerve fibers were more visible post-processing, providing a more uniform input for segmentation. Quantitative intensity metrics indicated a reduction in inter-slice variance[fig.11]. The detrended coefficient of variation decreased from 0.42 to 0.28 over the average intensities of the images, demonstrating improved intensity homogeneity.

4.2. *U-Net Feature Detection*

The 9-layer U-Net model demonstrated stable convergence over the full 50-epoch training process, achieving a Tversky-Focal Loss of 0.55 on the validation set. While the model achieved a high pixel-wise accuracy of 0.98, the Dice Coefficient of 0.29 (Dice Loss of 0.71) highlights the challenge of segmenting fine, sparse structures. This discrepancy is likely due to the extreme class imbalance, where the background dominates the loss calculation despite the use of Tversky-Focal weighting. Segmentation performance was strong for larger, contiguous structures, such as major blood vessels, which were consistently identified with well-defined boundaries across adjacent slices. However, smaller structures like capillaries, which are represented by small or even single-pixel values, were not well segmented. This limitation may be due to the loss function disproportionately prioritizing larger structures (comprising thousands of pixels) over smaller, sparse features. To address this, future work may introduce an additional class for capillaries, or focus on expanding the training dataset, as typical U-Net implementations for similar biological imaging tasks often utilize training sets of over 500 annotated images. In our study, we evaluated the method with a smaller set of 32 images, which likely contributed to the model's difficulty in segmenting fine structures. To partially mitigate this issue, we used the union of U-Net predictions with Difference of Medians masks to improve the segmentation of blood vessels.

4.3. *Difference of Medians*

The Difference of Medians approach worked well to remove noise and background features. Applying the filter decreased average image intensity from 11890 to 1054, indicating a decrease in the image noise floor and

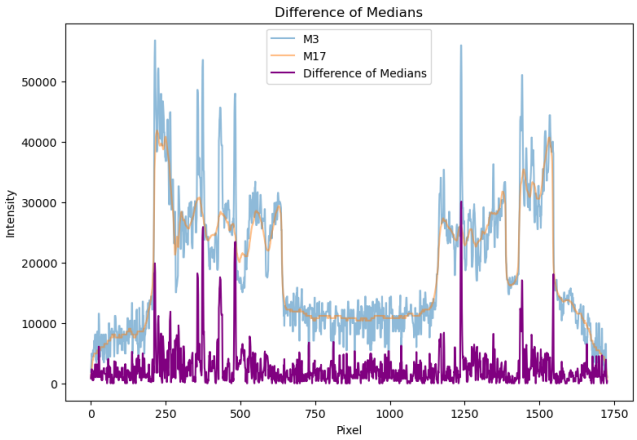

Figure 12. A one dimensional slice illustrating the medium-scale feature detection by the Difference of Medians approach.

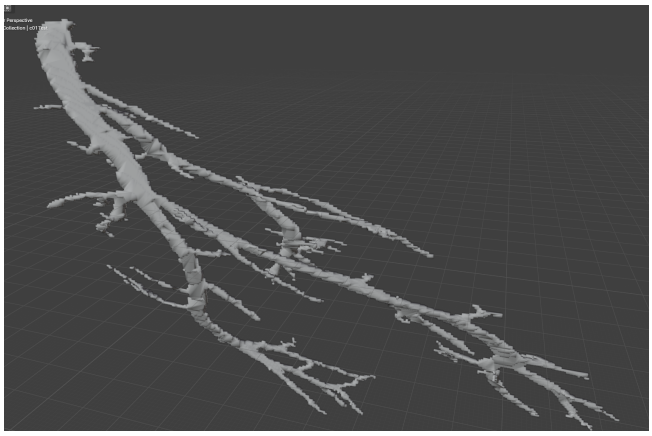

Figure 13. Connectivity analysis within the game engine. By selecting all topologically connected faces, a single vascular 'tree' can be isolated from the heart wall. This demonstrates the pipeline's ability to preserve structural continuity, enabling future pathfinding and transport modeling.

background structures. The average signal pixel value also went from 45672 to 16933, making a signal to noise increase from 3.8 to 16.0. Further, the signal peaks were both more well-defined and simpler to pick out with an intensity threshold. Though medium features were detected well, this method scales poorly to larger feature size band passes. With a time complexity of roughly $O(n \cdot m \cdot I^2 \log(I))$ Where I is the size of the large median kernel, this method becomes infeasible for large features. Additionally, the median kernel will become less effective at selecting larger features as the feature contributes more pixels to the median kernel.

5. APPLICATION AND FUTURE WORKS

The final models produced by this pipeline enable efficient visualization and analysis of the biological structures on standard computational hardware. By converting the segmented volumes into optimized mesh representations, we significantly reduce the computational footprint, allowing the models to be rendered and explored on a variety of platforms. Both models fit within a single game file of under 1 GB, demonstrating the feasibility of running such large-scale biological datasets on modest systems. Figure 13 shows a connectivity analysis that is simple to perform with data in this format, showing potential utilities that would be impossible with the original dataset.

Looking ahead, there are several avenues for further development and refinement. One simple improvement could be the implementation of a blind point spread deconvolution. Algorithms such as the Richardson Lucy transform have been modified to both derive a pseudo-psf and operate with a varying psf across an image. The U-Net model could include separate segmentation classes of larger vessels and smaller capillaries within the model. Given the challenges faced in segmenting smaller structures, a separate class for small features could reduce the impact of lopsided data. Additionally, increasing the size of the training dataset could improve model performance. Given that current U-Net implementations for similar biological imaging tasks often rely on training sets with over 500 annotated images, a larger training dataset would likely yield better generalization and segmentation of small-scale features. There is also significant work on 3D U-Net implementations, which could leverage the z continuity of these structures for improved performance at the cost of a more complicated training data preparation and a higher memory cost in training.

Furthermore, while the current dilation/erosion pipeline for postprocessing is effective, it remains a relatively simplistic method for bridging voxelized structures. There is significant room for improvement in this area, as more sophisticated algorithms for voxel connectivity and structure bridging exist in graph-based or morphological techniques. These algorithms are being applied on similar datasets like neuron tracing to help improve the continuity and accuracy of segmented structures, particularly in regions where adjacent structures are sparsely connected across slices.

In summary, while the presented pipeline provides a functional and efficient means of segmenting and visualizing large-scale biological datasets on modest computational resources, further optimization and refinement of segmentation models, training sets, and postprocessing techniques will be necessary to address the challenges of segmenting smaller, more intricate structures and fur-

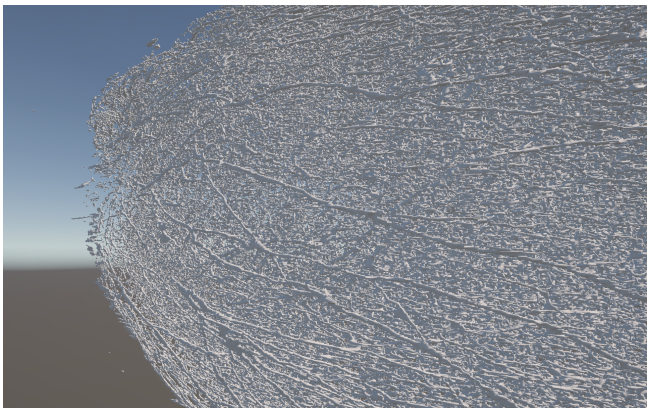

Figure 14. A screenshot of the nerve structure as it appears in the game engine.

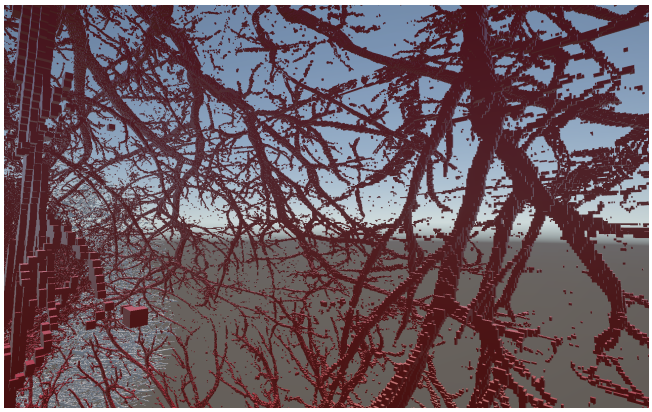

Figure 15. An image from within the heart of the blood vessel channel.

ther enhance the utility of this method for biological research.

6. DATA AND CODE

- GitHub for Data Pipeline - https://github.com/Thomaslund1/Heart-Scan-Pipeline.git
- GitHub for Demo Render Shown - https://github.com/Thomaslund1/Heart-Scan-Unity-Demo.git
- Home of the Human Protein Atlas - https://www.proteinatlas.org/

REFERENCES

Lorensen, W. E., & Cline, H. E. 1987, ACM SIGGRAPH Computer Graphics, 21, 163, doi: 10.1145/37402.37422

Ronneberger, O., Fischer, P., & Brox, T. 2015, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, ed. N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Cham: Springer International Publishing), 234–241

Uhlén, M., Fagerberg, L., Hallström, B. M., et al. 2015, Science (New York, N.Y.), 347, 1260419, doi: 10.1126/science.1260419

Zekri, M. A., & Lang, I. 2024, Methods in molecular biology (Clifton, N.J.), 2772, 353, doi: 10.1007/978-1-0716-3710-4_27

Zuiderveld, K. 1994, in Graphics Gems IV (USA: Academic Press Professional, Inc.), 474–485