

Semi-Automated Technique for Extraction and Segmentation of Neuronal Cell Bodies and Nuclei

Peter Chen^{1,5}, Maher Shammaa², Michael Morehead³, George Spiro⁴

¹Summer Undergraduate Research Internship (SURI) Program, ²Department of Chemistry,

³Lane Department of Computer Science and Electrical Engineering, ⁴Center for Neuroscience, West Virginia University School of Medicine

⁵Department of Computer Science, Princeton University

Problem Statement

In connectomics, the mapping of the brain's synaptic connections and wiring, segmentation requires extensive time and manual effort. Automation of various pipelines in segmentation can dramatically reduce the level of effort required to analyze connectomic data. One of the greatest challenges is the visualization of the structure of the nervous system at the resolution needed to discern neural connectivity. Thanks to automated serial electron microscopy techniques and computational advances, it has recently become possible to acquire image data from nervous tissue at an unprecedented rate.

However, computers store these data and require programming in order to manipulate the data in a meaningful way. Humans, by contrast, are adept at contextualizing regions of data to identify cell bodies, nuclei, axons, and synaptic inputs within the image volume. Manual work in the effort to segment images is slow and expensive due to the enormous size of the datasets, and automation is challenging due to the highly convoluted nature of nervous tissue.

Hypothesis

By using a novel semi-automated method for the extraction of neuronal cell bodies and nuclei we can dramatically decrease the time needed to segment cell bodies from raw EM data because you can exclude a large fraction of the image volume that does not contain these objects (fig.1).

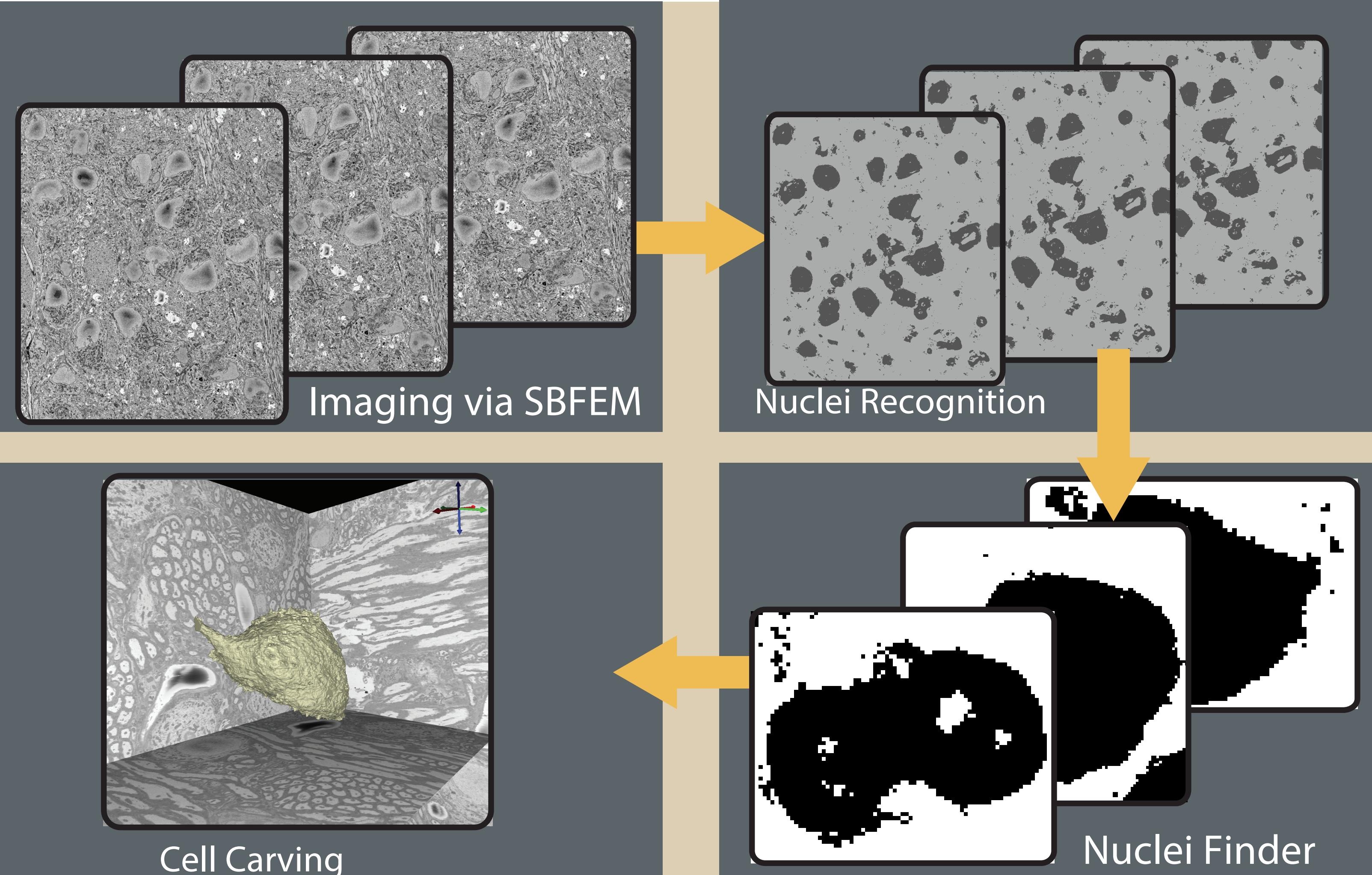


Fig. 1. Clockwise, the pipeline from start to finish of our semi-automated approach to finding and carving cells. 1. Imaging via Serial Block-Face Electron Microscopy, 2. Nucleus recognition with Ilastik. 3. Nucleus Finding and Cropping with our program. 4. Cell carving with Ilastik.

Methodology

Pixel Classification by ILASTIK (fig.3a)

1. Input EM data into Ilastik Machine learning program

i. Uses the machine learning pixel classification to perform a simple segmentation of the cell nuclei,

ii. nuclei are more uniform in shape and pixel intensity than cell bodies

Nucleus Finder (Proprietary code; fig.3c.)

3. Output is then fed into a Nucleus Finder. Calculate the centers of these nuclei. (fig.1b.)

i. cycle through each slice in the simple segmentation data and find all shapes that are large enough to be nuclei

iii. output data of all nuclei found

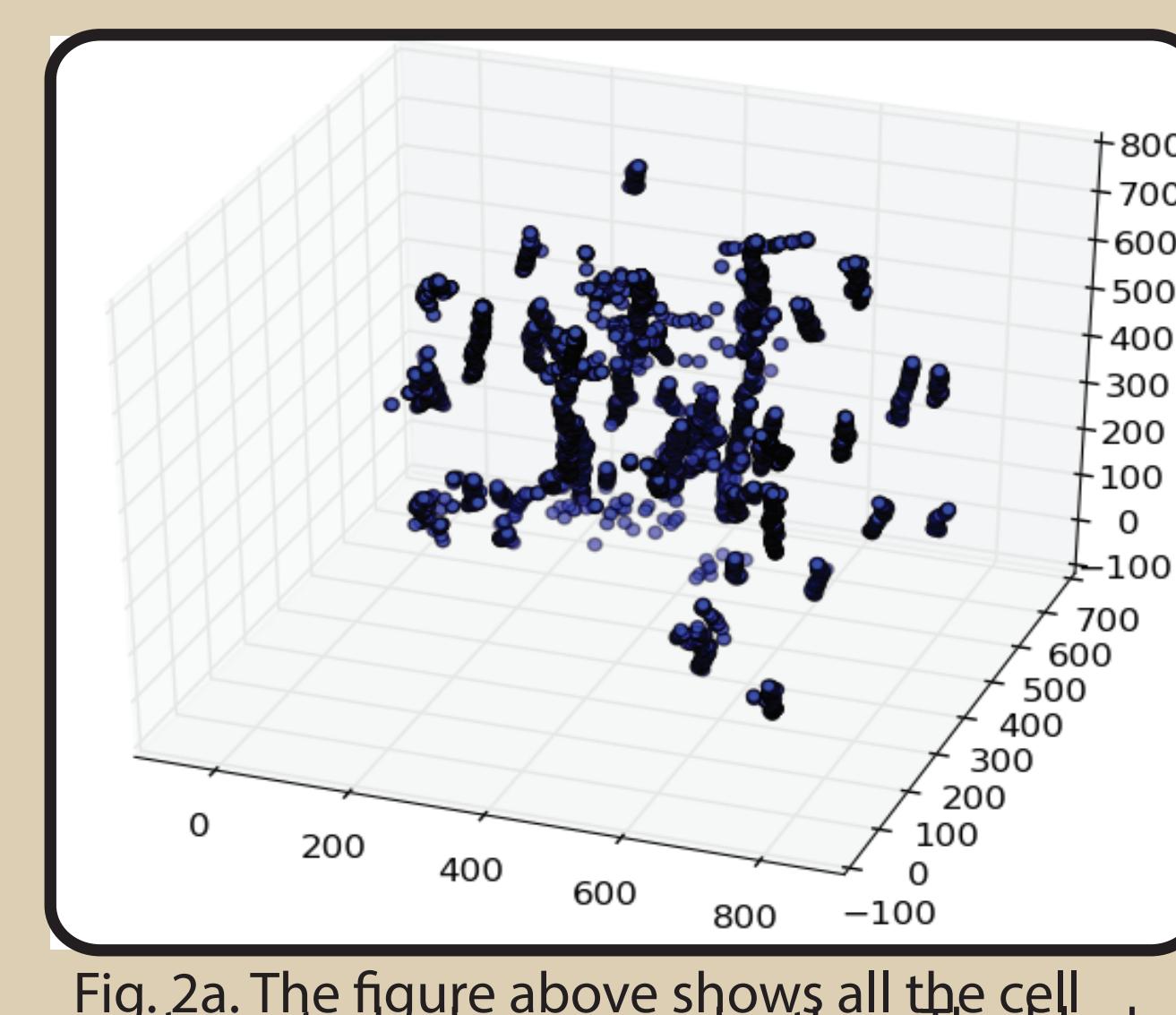


Fig. 2a. The figure above shows all the cell centers, stacked up on each other. The black columns are the cells, and the blue smaller circles are false positives.

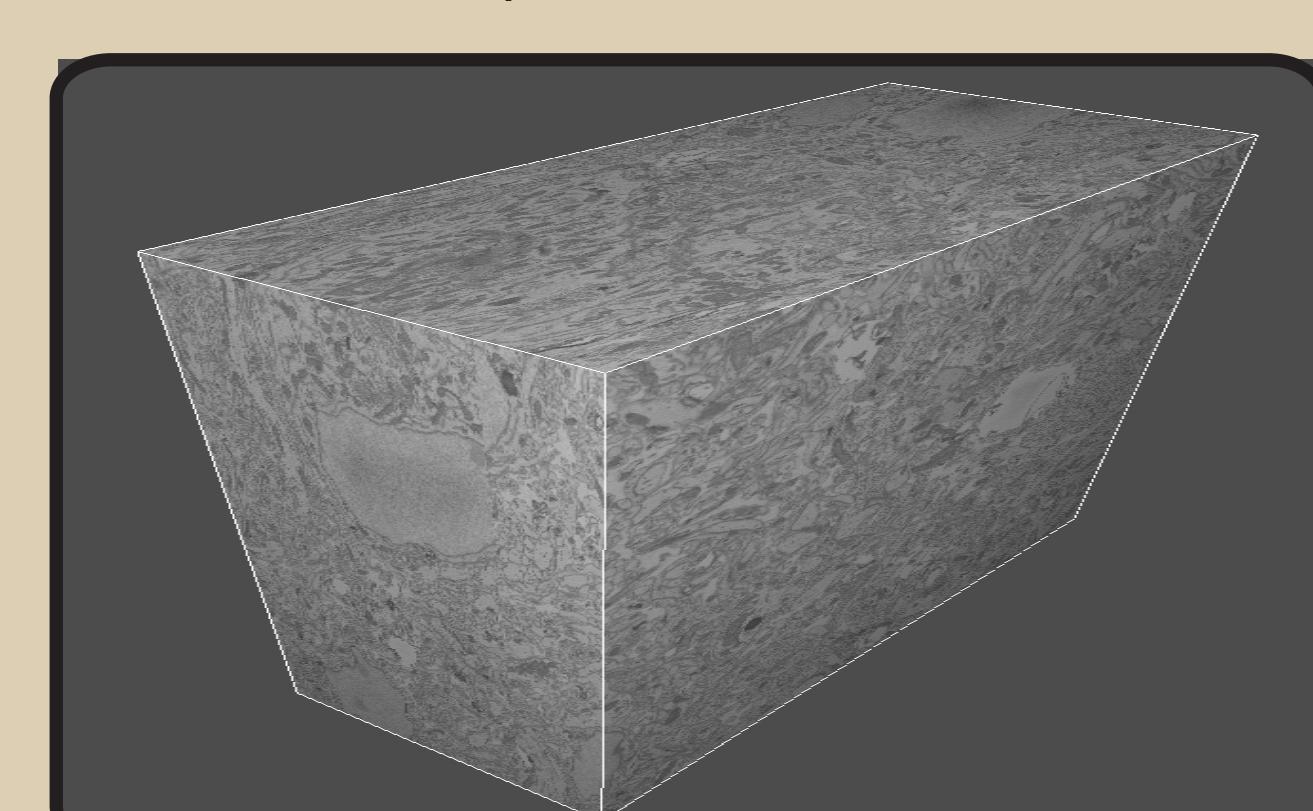


Fig. 2b. The result of the Nuclei Cropper crops out a rectangular prism for each cell body in the original EM data at the original resolution.

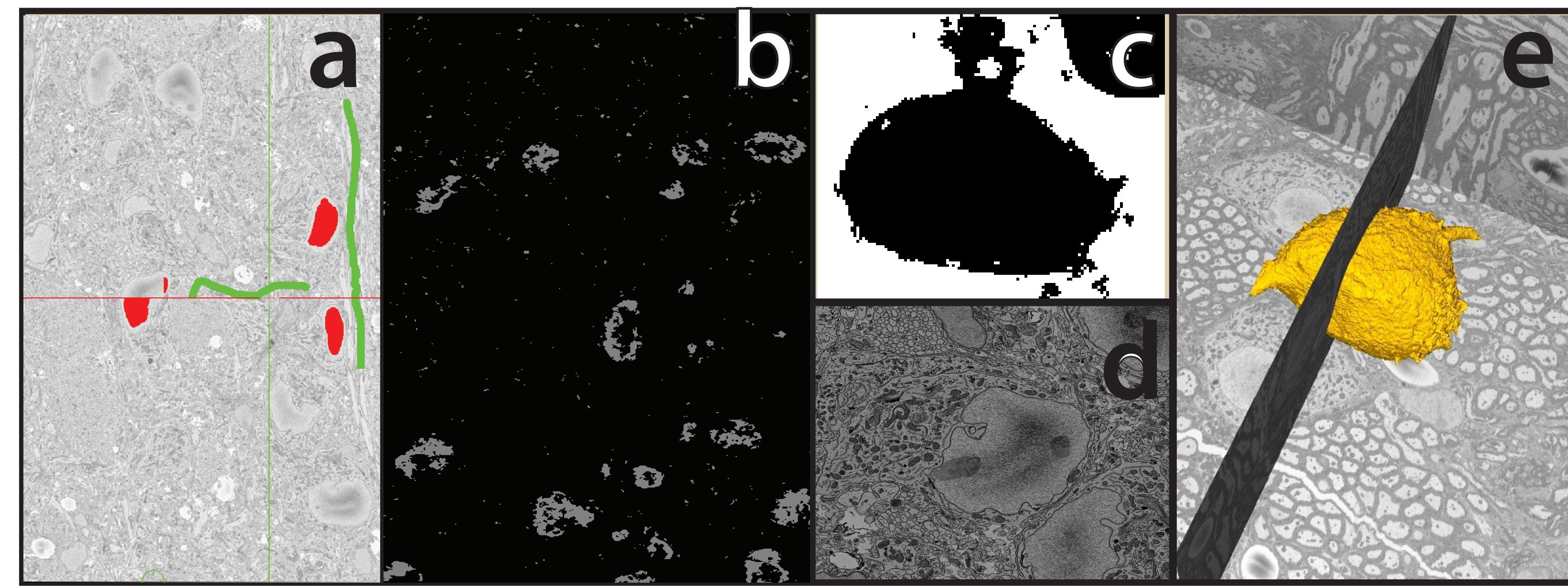


Fig 3. a. Ilastik Training. b. Ilastik output. c. Nucleus Finder d. Cropper output. e. Cell body fully carved

Methodology (cont.)

Nucleus Cropper (Proprietary code; fig.3d.)

4. Create appropriately sized cropping boxes, generating a subvolume for each nucleus large enough to encapsulate the associated cell body. (fig.2a.)

Cell Body Carving by ILASTIK and VIGRA (fig.3e.)

5. Once a subvolume has been generated for each cell, they can be fed into Ilastik and carved. VIGRA is a preprocessing program that applies filters to make carving easier

Results

The confusion matrix below shows our preliminary results for the classification of cells from our program and compared test data from a human. There are 37 verified cells in post-natal day 3 data. Our program found 27 of the 37 cells. The program also found 26 false positives. These false positives were most likely clusters of mitochondria.

Approximately 45 hours of time spent manually cropping, and segmenting cell bodies now only takes less than 20 minutes.

		Confusion Matrix for Cell Finder	
		True Neuron	False Neuron
Actual Class	True Neuron	0.579	0.175
	False Neuron	0.246	0.0
	Predicted class	True Neuron	False Neuron

Fig. 4a. The Confusion Matrix showing the False Positive, True Positive, True Negative, and False Negative.

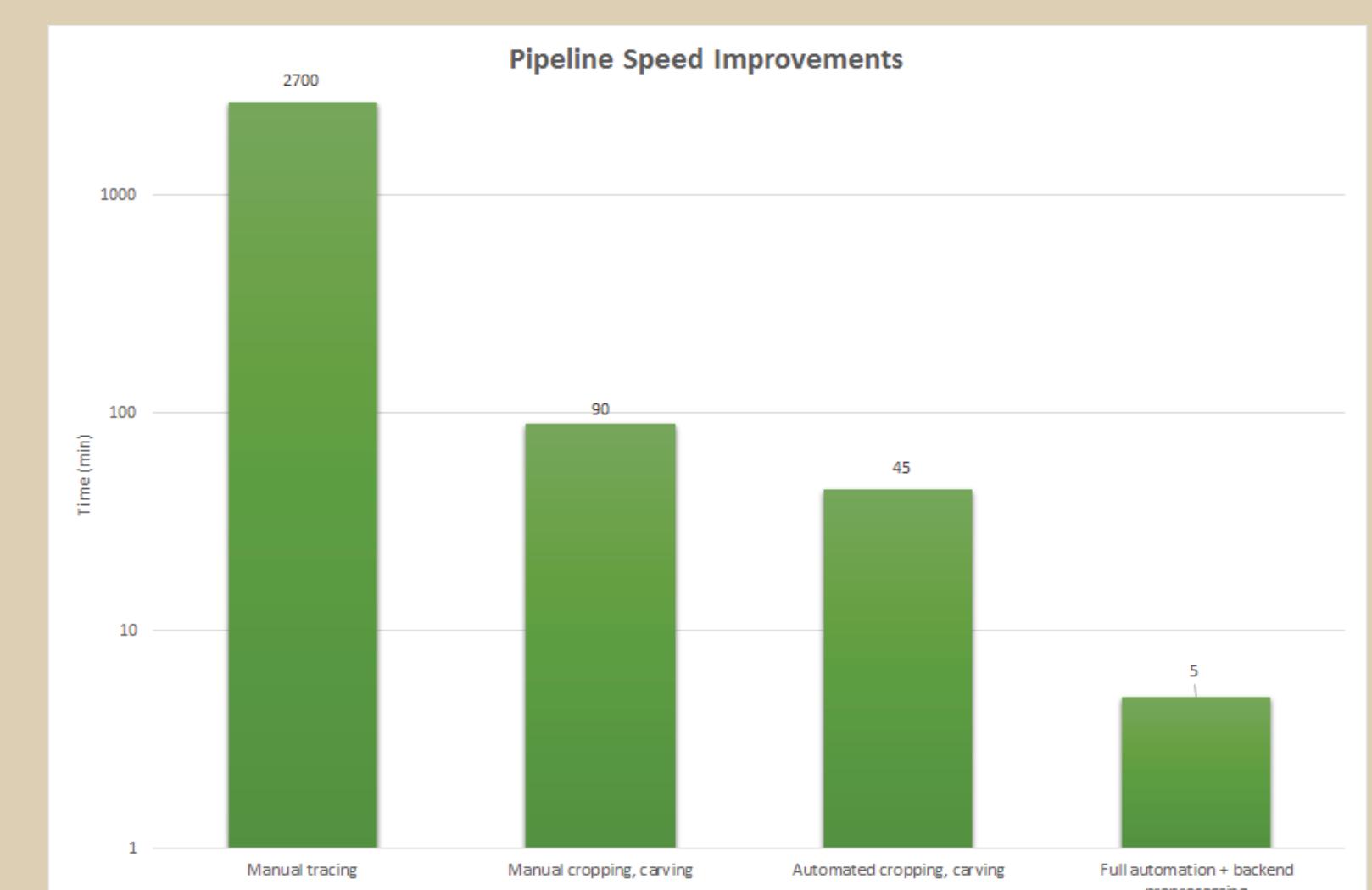


Fig. 4b. A bar graph showing the decrease in time from the manual carving compared to our result.

Discussion

The resulting confusion matrix demonstrates our hypothesis that the use of a semi-automated workflow can dramatically increase efficiency. This was possible due to removing the human factor from the pipeline by and removing a large fraction of the data, focusing only on cell bodies.

Future Work

Given more time, we would like to perfect our program by:

1. Acquire more testing heuristics including: larger test cases, difficult test cases (irregular data).
2. Outputting files already converted for pipeline processing.
3. Preprocessing of Ilastik using Ilastik headless mode.
4. Port Python scripts to Scala or C: decrease time, decrease memory usage.
5. Develop more accurate Nuclei Finder to rule out false positives.

Acknowledgements

I would like to thank Michael Morehead for guiding my progress. Maher Shammaa for assisting me with programming. The rest of the lab and crew for their support. Dr.Spiro for accepting me into the SURI program.