Mapping the Brain: An Introduction to Connectomics Streamlining the NeuroData MACHO Pipeline

Thomas Keady, Albert Lee, Augusto Ramirez

January 9, 2016

1   Introduction

Creating a complete wiring diagram of neuronal connectivity is one of the most ambitious goals of connectomics and contemporary neuroscience [1]. Thanks to technical and computational advances that automate the collection of high resolution electron-microscopy data, complex networks of neurons and their synaptic connections can be directly observed and eventually fully mapped. NeuroData's Open Connectome Project (OCP) includes Machine Annotation for OCP (MACHO) which is one of the image preprocessing tools developed to make the automated images-to-graphs (i2g) pipeline more efficient and compatible. While a number of machine annotation tools exist, only MACHO allows for efficient integration with the i2g pipeline. However, i2g's reliability depends on pre-pipeline image processing from algorithms such as watershed and gala found in MACHO. Issues arise from noise generated by image processing which can result in significant errors in output graphs. Currently, performance error rate ranges between 65% to 70%.

Vectorization of the EM images for pre-pipeline processing can cause, among other things, breaks in the continuity of membranes. A very small error in this step can result in significant error carried forward to the graph. Cells that have a membrane between them can become connected (and vice-versa, a single cell can be split into two or more). Code exists for gap completion, which can fix errors like these. We plan to introduce this gap completion code as another step of preprocessing before the i2g pipeline. If done properly, we anticipate that this additional step can prevent significant amplification of error in the i2g pipeline by fixing broken membranes in early processing stages. This would increase the accuracy of the output graph compared to "ground truth".

2   Project Outline

2.1   Schedule

- 1/9/16: Talk to Will and Greg about where we can actually run our modifications and the whole pipeline. Finalize and locate the program set-up and analysis tools/resources we will be using.

- 1/12/16: Compare and contrast at least two sets of images through our pipeline and the current pipeline and most importantly, understand the results.

- 1/14/16: Have Matlab segments of the pipeline converted to working and integrated Python.

- 1/16/16: Begin implementing Kaynig gap completion code in Python as another step in the pre-pipeline processing.

- 1/22/16: Be able to run full pipeline plus our addition(s).

- End of class: Further reduce inaccuracies.

2.2   Methods

Using code from the MACHO pipeline we will modify the algorithms currently used in order to improve accuracy. Specifically we will start with the watershed pre-processing and use Kaynig's gap-closing program. Additionally we will try using Gala pre-processing and post-processing and see if that creates an impact. Finally regardless of the modifications we create we will streamline the pipeline by converting the necessary matlab files into python files.

2.3   Allocation of Tasks

Albert Lee: Project research, documentation, algorithm comparison
Thomas Keady: Algorithm writing, documentation, Matlab-to-Python
Ausuto Ramirez: Data analysis, documentation, graphics

References

[1] W. G. Roncal, D. M. Kleissas, J. T. Vogelstein, "An automated images-to-graphs pipeline for high resolution connectomics." *Frontiers in Neuroinformatics*, vol. 9, no. 20, Aug. 2015.