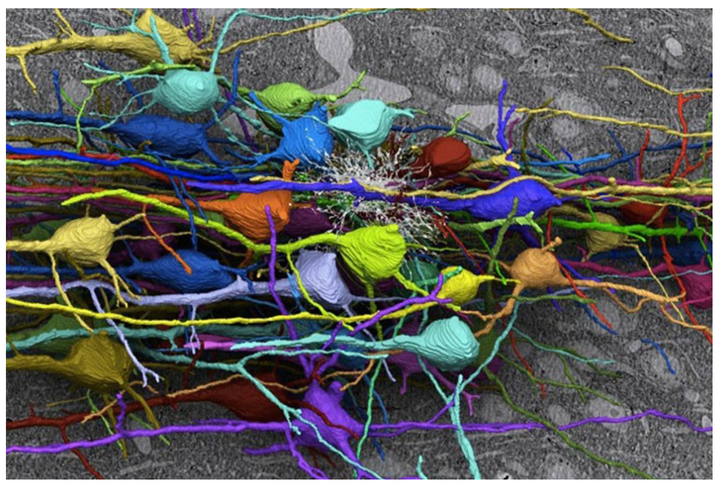
Big Data Agglomeration for Connectomics

Massachusetts Institute of Technology
School of Engineering
SuperUROP
2015–2016

Hayk Saribekyan, Yaron Meirovitch, Nir Shavit (MIT EECS)
MIT EECS – Landsman Undergraduate Research and Innovation Scholar

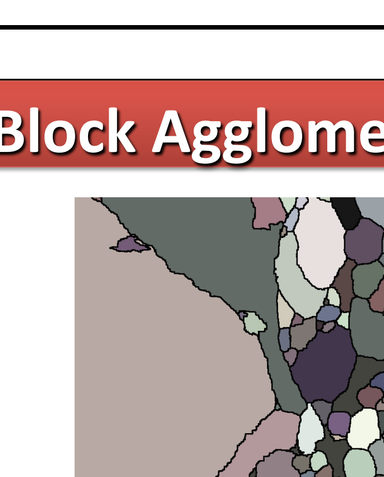
Background: What is Connectomics?

- Connectomics aims to produce a dense mapping of cells and their connections in the nervous system of an organism
- Current Research is focused on segmentation of 3D electron microscopy images into individual cells



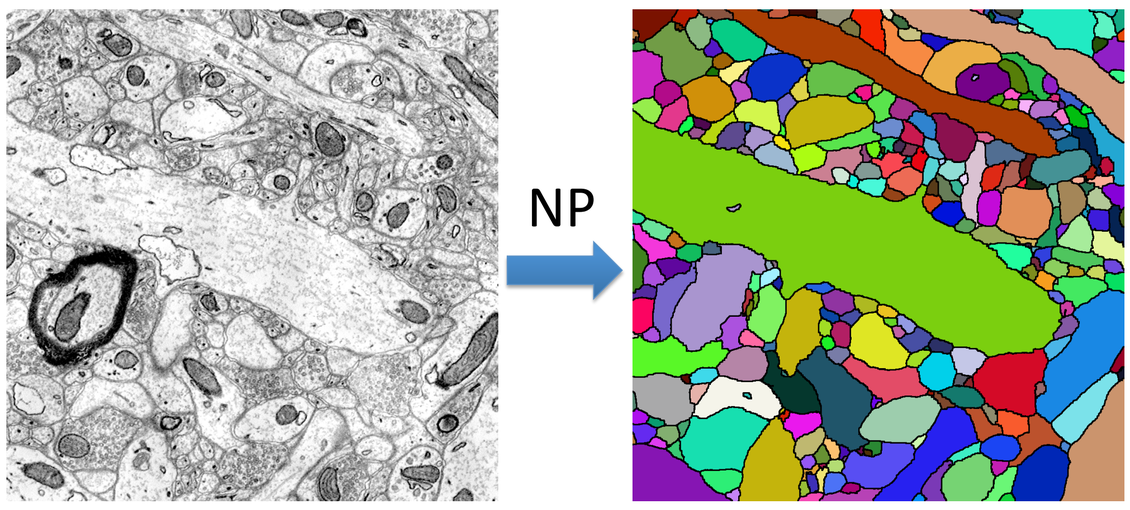
Few segmented cells in mouse cortex
Kasthuri et al.

High Performance Segmentation Pipeline



- Images correspond to 30nm slices of the tissue
- Data representing 1 cubic millimeter of tissue is 2PB!
- The pipeline operates on a **single 72-core** machine for high parallelism

Segmentation of Blocks



EM image Segmentation

- NeuroProof (NP) generates high quality segmentation for GB sized blocks
- Does not scale to larger datasets
- Need to agglomerate segmentation from multiple blocks

Block Agglomeration

- A small sub-volume of blocks is enough to decide merges of segments between them
- Use NeuroProof again on the sub-volume to detect pairs of segments to merge
- Disjoint-set data structure for efficient relabeling of segments

Results and Evaluation

- Less than 10 hours to process 1TB data
- State-of-the-art quality of segmentation
- Achieves VI metric of about 1.66 for areas that have ground truth

- Automatically constructed from data by Kasthuri et al.
- Spans more than 100 segmentation blocks
- Traverses about 40µm in space

Next Steps

- Reconstruction of neural connectivity graphs from cell skeletons and synapses
- Building efficient data structures to query the connectivity graphs and to merge them at a larger scale
- Continue improving the quality of segmentation

References

[1] Saturated Reconstruction of a Volume Neocortex, Kasthuri et al., 2015
[2] https://www.janelia.org/open-science/neuroproof
[3] Variation of Information, Melia M., 2003