

NEAR REAL-TIME PROCESSING OF LIGHT-SOURCE WORKLOADS AT ALCF

Experiment

1. Transfer 40GB+ input dataset from APS, ALS, and NSLS-II
2. Process data with real-time queue
3. Transfer results to

Results

- Continuously ~~executed~~ *and simultaneously* for 48+ hours
- Transferred 23TB input data from APS/ALS/NSLS-II to ALCF (average bandwidth: 276Mbps)

Toward Real-time Analysis of Experimental Science Workloads on Geographically Distributed Supercomputers

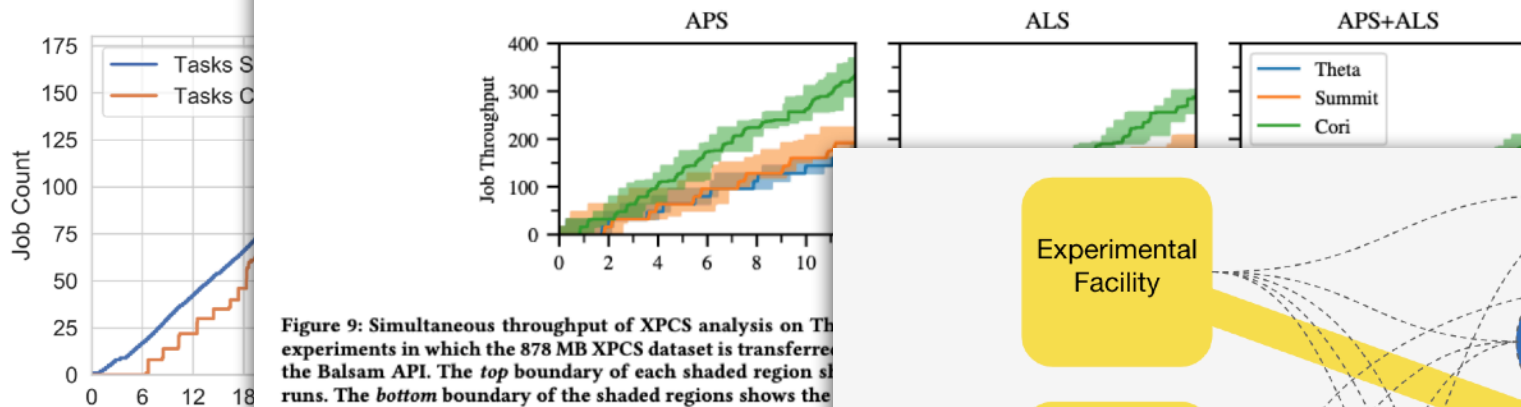
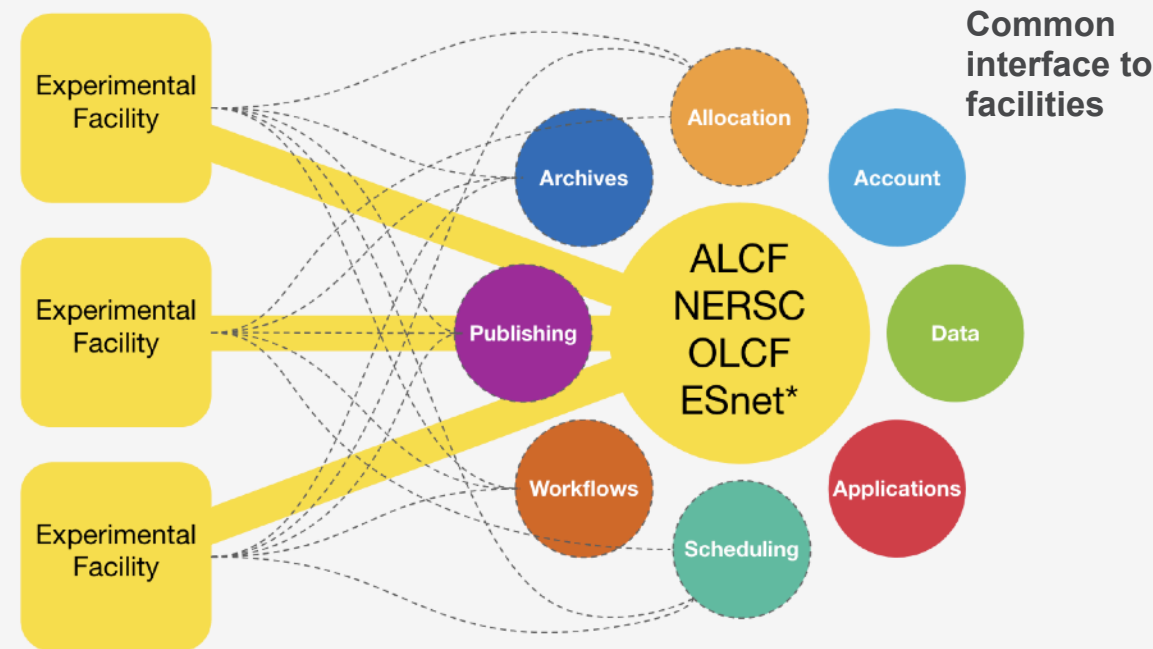


Figure 9: Simultaneous throughput of XPCS analysis on the experiments in which the 878 MB XPCS dataset is transferred via the Balsam API. The top boundary of each shaded region shows the system is network I/O-bound: that is, compute nodes become

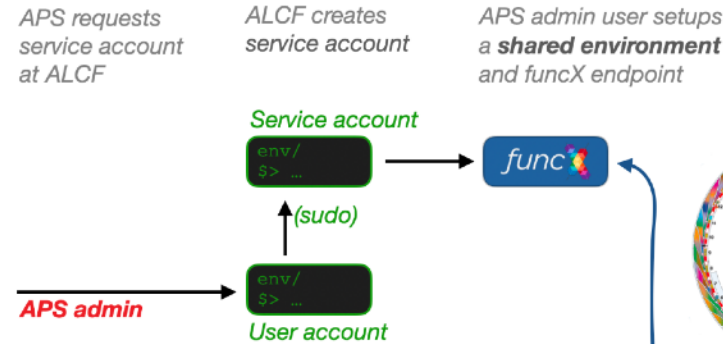


Nexus@ALCF

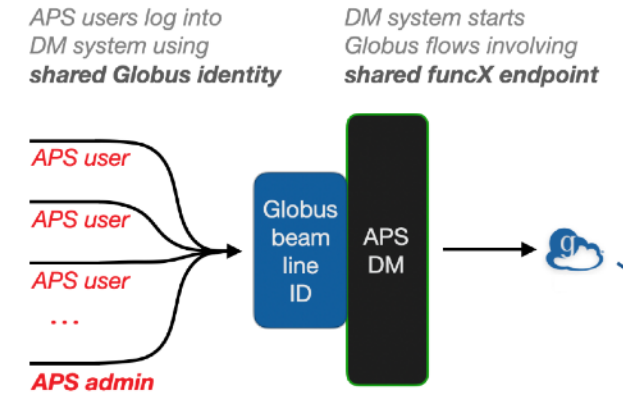
1. Instrument accounts to provide identities
2. Demand and preemptable queues enable on-demand access
3. Allocations and suballocations empower facilities to manage resources
4. Fully automated

One-time configuration at ALCF | APS experiments

Using a service account per beam line



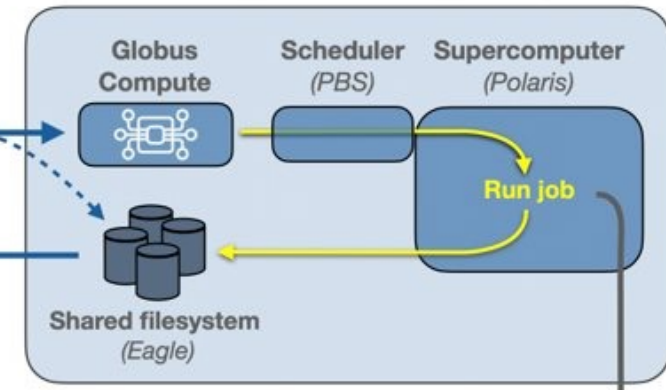
No human involved in the workflows



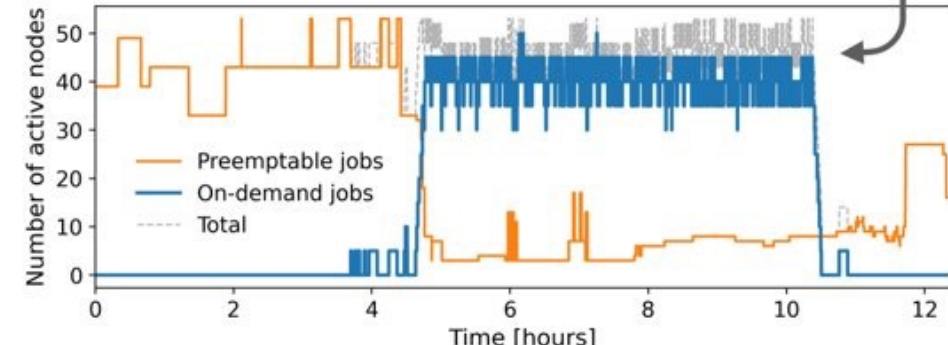
APS (data acquisition)



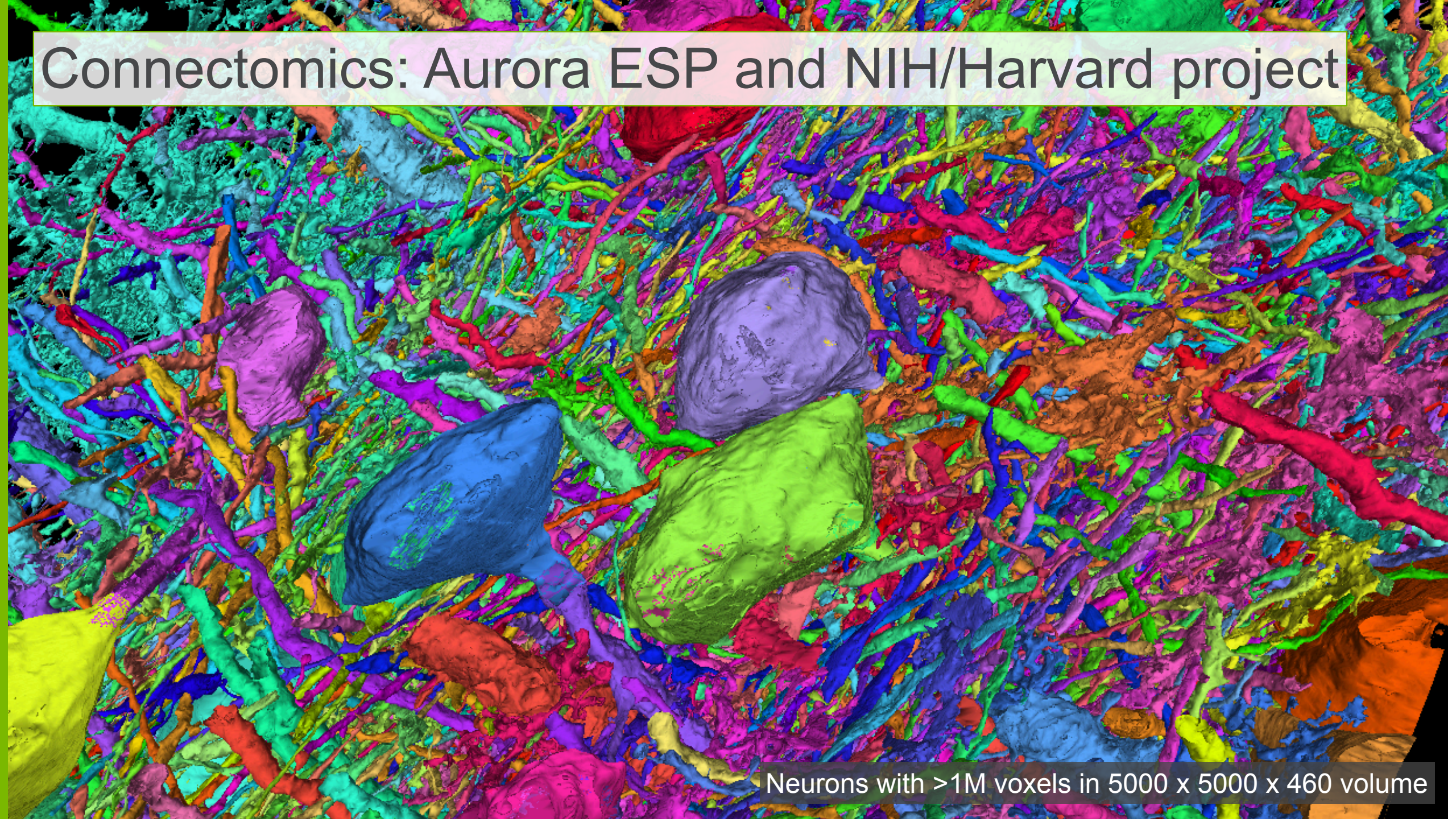
ALCF (on-demand data analysis)



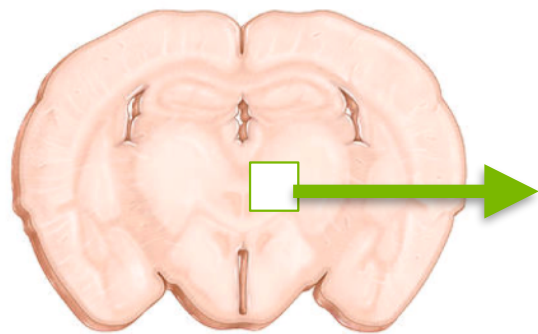
Computation example (Part of Laue experiment) April 1st, 2023



Connectomics: Aurora ESP and NIH/Harvard project



Neurons with >1M voxels in 5000 x 5000 x 460 volume



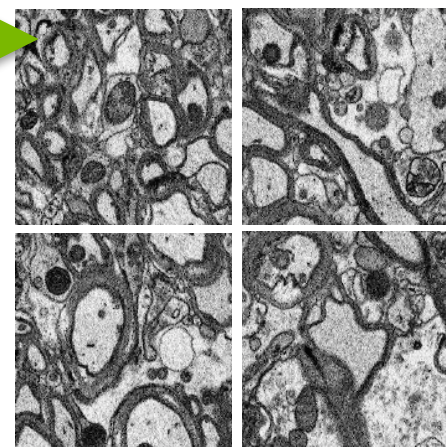
Mouse brain:
70M neurons
~1cm³



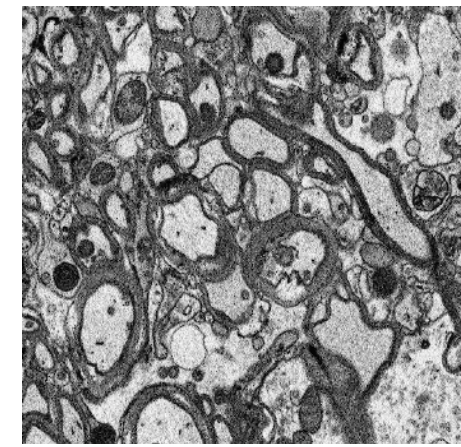
~1mm³



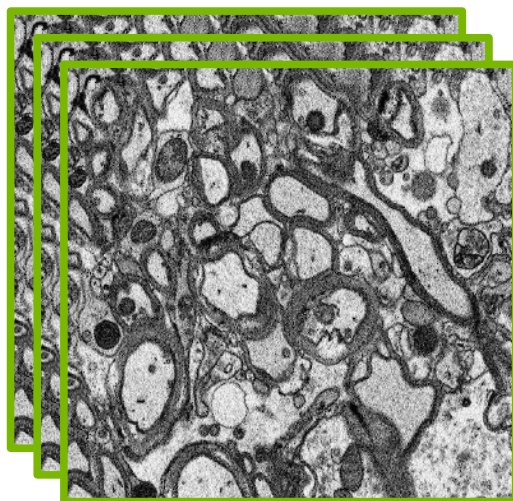
25000
40nm sections
1mm x 1mm
(4nm resolution)



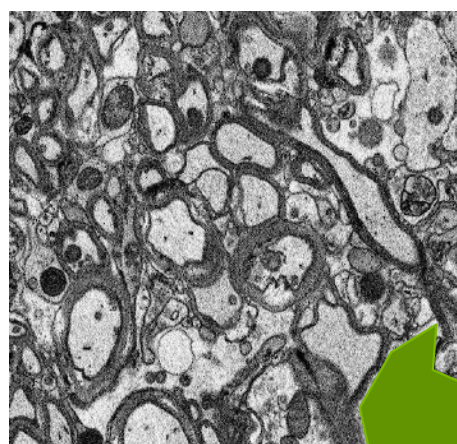
Each section
imaged with EM as
N tiles (8 bit)



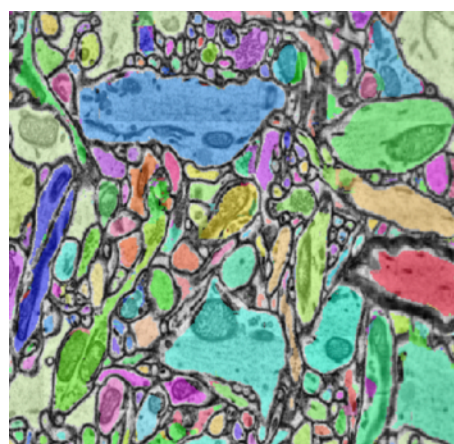
Sections stitched
together (CV-based)



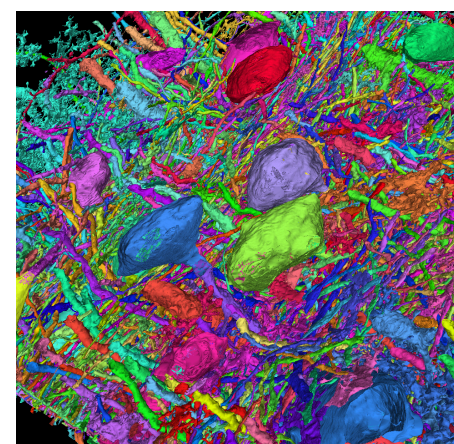
Aligned sections
(CV-based)



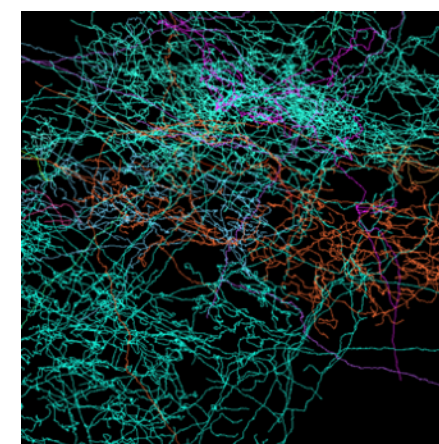
Mask out
non-target
objects



Segment
target
objects



Reconstructed 3D
neurons with >1M
voxels
(NN-based)



Skeletonize to
produce connectivity
graph for analysis