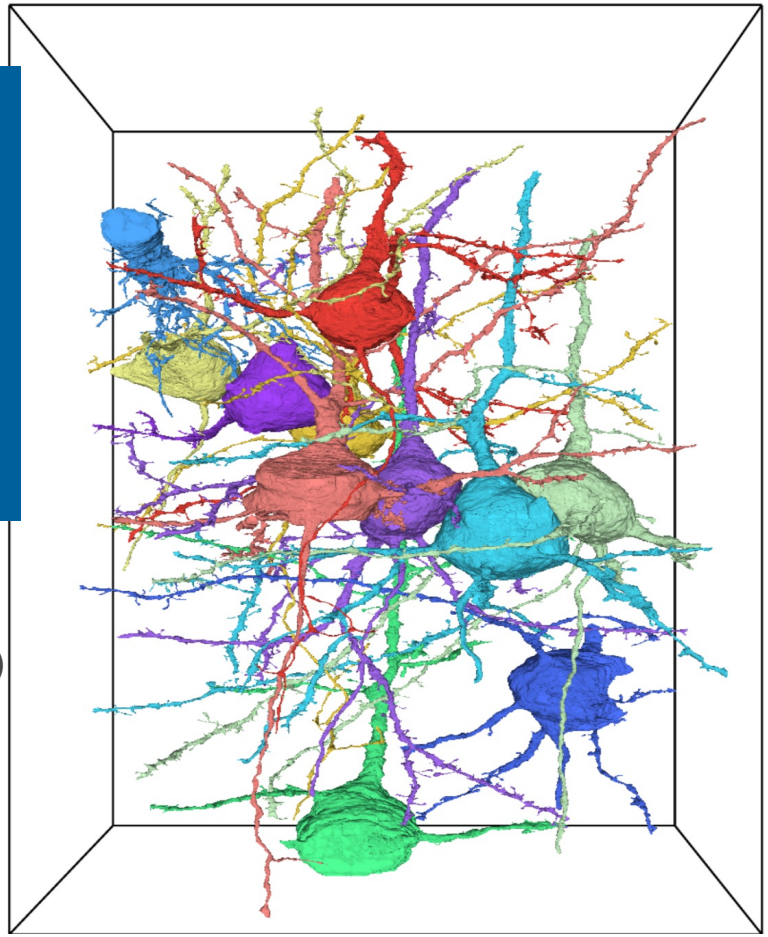
ROBUST SCIENCE WORKSHOP

FEBRUARY 17, 2021

Argonne NATIONAL LABORATORY

TOWARDS AN HPC AUTOMATED PIPELINE FOR CONNECTOMICS



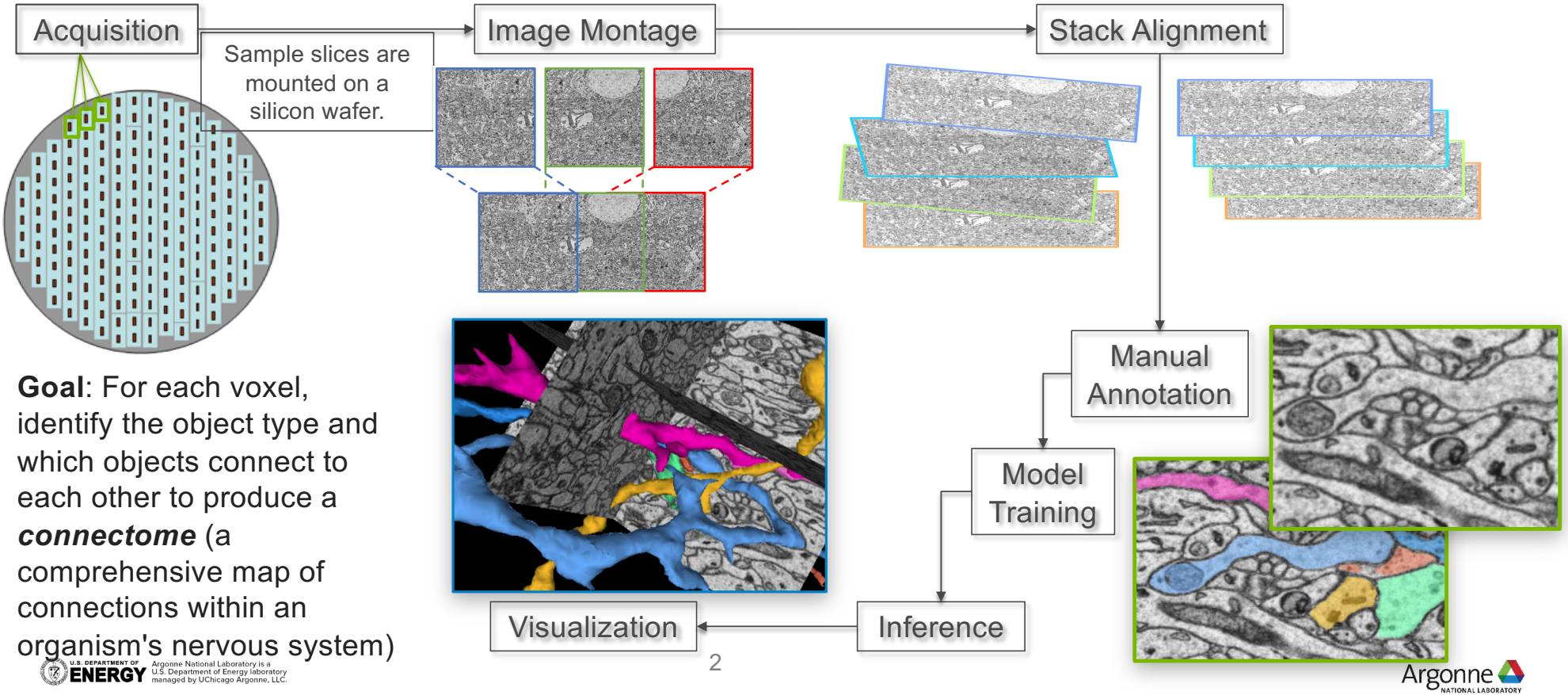
NICOLA FERRIER, ARGONNE NATIONAL LABORATORY

Rafael Vescovi, Murat Keceli, Misha Salim, Thomas Uram (ANL)
Hanyu Li, Narayanam Kasthuri (Univ of Chicago)

U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

02/17/2021

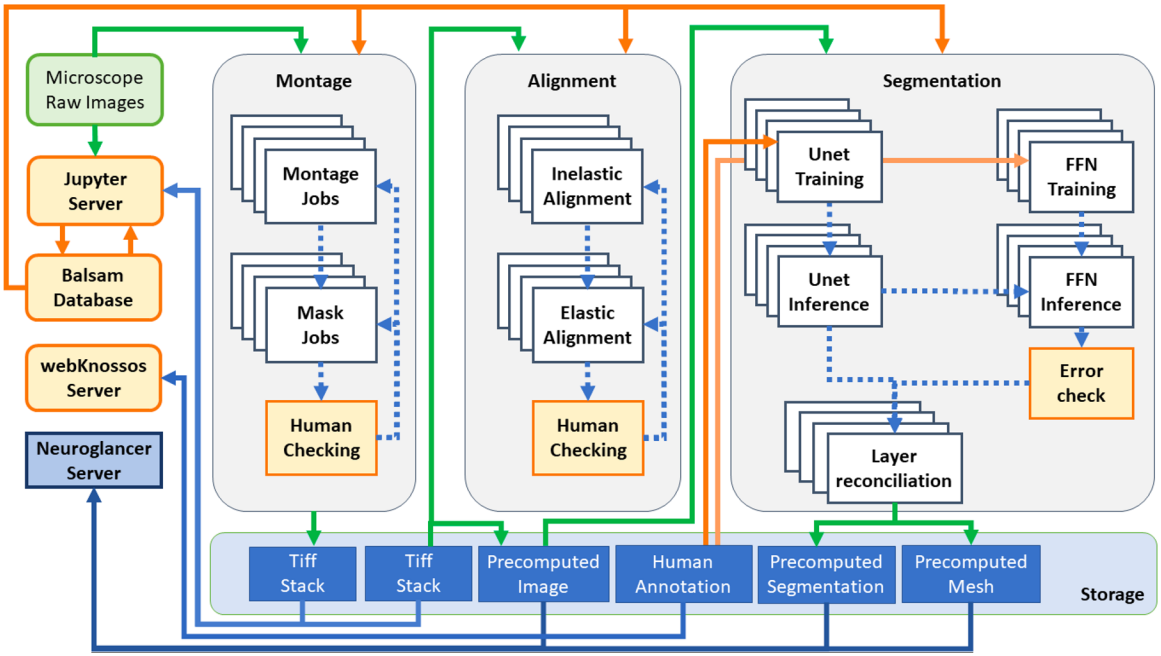
SERIAL ELECTRON MICROSCOPY ACQUISITION AND ANALYSIS STEPS FOR CONNECTOMICS



Goal: For each voxel, identify the object type and which objects connect to each other to produce a ***connectome*** (a comprehensive map of connections within an organism's nervous system)

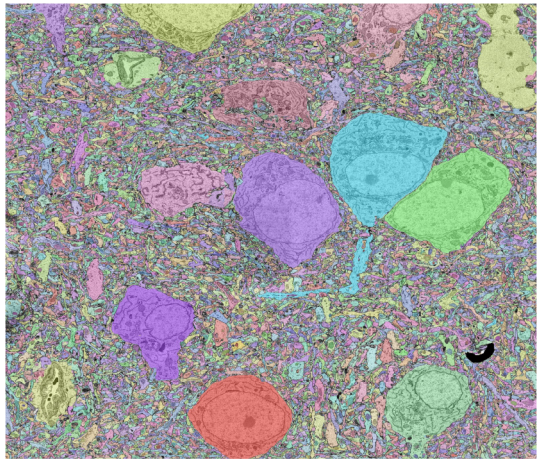
Scalable, trustworthy, and reproducible
Our experience: EM PIPELINE @ ARGONNE

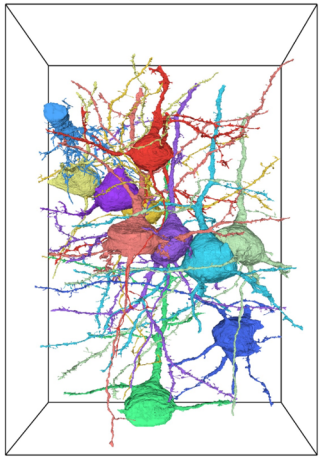
- **Optimizing codes** for montage, alignment, and segmentation to run on Argonne supercomputers with natural concurrency (tens to thousands of compute nodes)
- Assembling codes into a pipeline
  - Respect (and exploit) concurrency of individual applications
  - **Reduce human-intensive aspects of large-scale computing**
  - Schedule jobs in a more optimal/efficient fashion
- Support user interaction with data throughout pipeline (orange boxes)

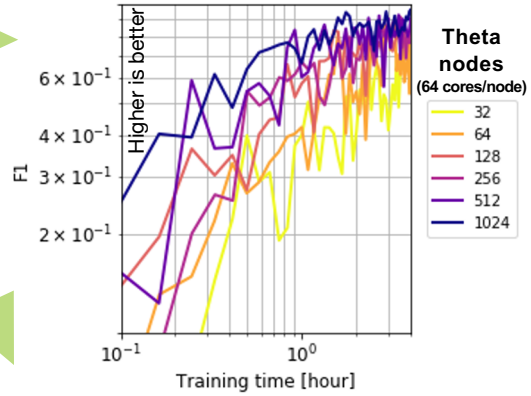


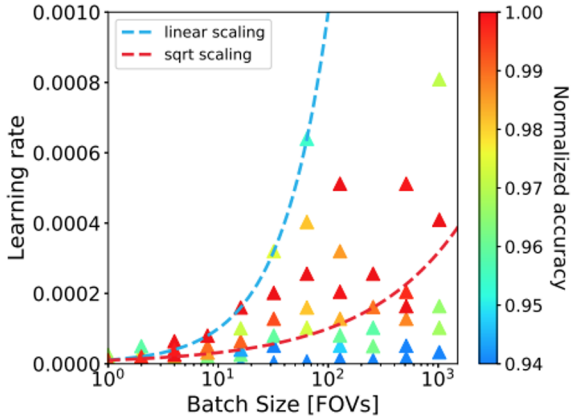
LARGE-SCALE FLOOD FILLING NETWORKS (FFN)

- **Scaled training of Google's FFN** to thousands of nodes on ALCF's Theta
  - Large-scale training achieves greater accuracy faster
  - Faster training opens opportunity to make many runs to reach optimal model (hyperparameter optimization)
- **FFN Segmentation** rate on Theta: ~350M voxels/nodehour
  - Distributed inference on many nodes allows us to reconstruct larger volumes faster (weak scaling)









Wushi Dong, Murat Keceli, Rafael Vescovi, Hanyu Li, Corey Adams, Tom Uram, Venkatram Vishwanath, Bobby Kasthuri, Nicola Ferrier, Peter Littlewood, "Scaling Distributed Training of Flood-Filling Networks on HPC Infrastructure for Brain Mapping", 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS) at SC19

WORKSHOP QUESTION

Can publishing/dissemination/sharing standard practices and procedures contribute to achieving or encouraging scalability, trust, and reproducibility in the applications results?

- Neuroscience computational community shares procedures, practices, data
- Scalability being addressed by *some*
- Trust – from human evaluation – bottleneck!

Main challenges:

- Reducing human effort (proof reading)
  - Develop better metrics for intermediate steps

ACKNOWLEDGEMENTS

Nicola Ferrier
Murat Keceli
Misha Salim
Tom Uram
Rafael Vescovi

Wushi Dong
Bobby Kasthuri
Hanyu Li
Peter Littlewood
Vandana Sampathkumar
Gregg Wildenberg

Jeffrey Kinnison
Walter Scheirer

Viren Jain
Michal Januzewski

Sebastian Seung's Lab

Supported by Argonne's Leadership Computing Facility
Aurora Early Science Program