Integrating Computer Vision into Microscopy Analytics Workflows

Matthew Guay [1]

[1]Laboratory of Cellular Imaging and Macromolecular Biophysics. NIBIB, NIH

February 14, 2019 ♥

National Institute of Biomedical Imaging and Bioengineering

MICROSCOPY AT THE NIH

NIH labs use microscopy to investigate biological structures from whole organisms to molecules.

We develop sophisticated workflows to turn data into clinical and fundamental biological insights.

New hardware technology gives us access to ever-growing datasets, but how can we harness them?

MICROSCOPY RESEARCH WORKFLOWS

This talk focuses on microscopy research workflows that look like:

- Image acquisition
- Image preprocessing (alignment, denoising, deconvolution, etc.)
- Image processing (segmentation, detection, tracking, etc.)
- Image analysis and visualization.

COMPUTER VISION FOR MICROSCOPY

Currently there is interest in computer vision to automate image processing problems in microscopy.

Computer vision (CV): Using algorithms to extract information from images "like humans do".

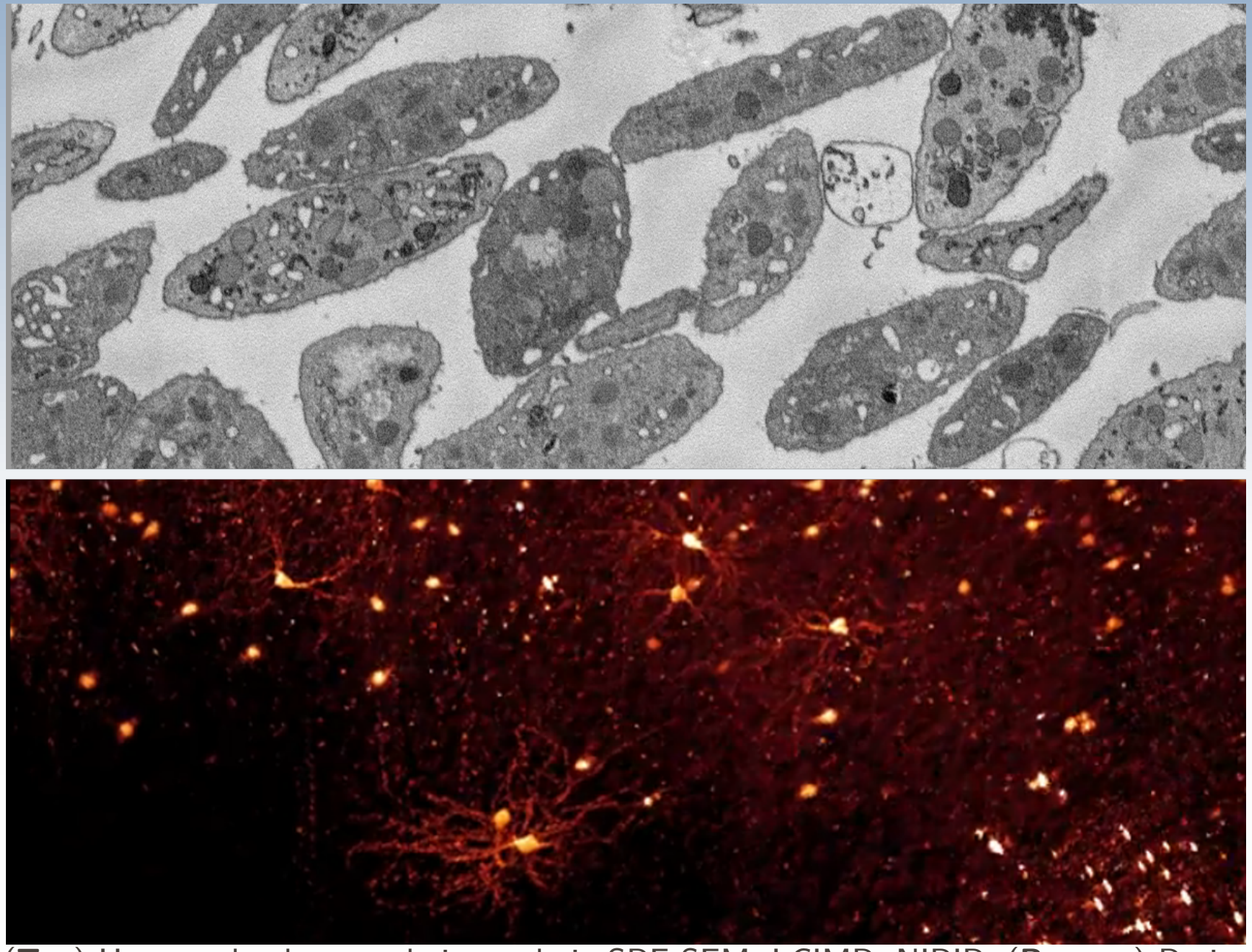
New machine learning techniques are improving CV performance, but accuracy and scaling challenges remain.

SCALING MICROSCOPY WORKFLOWS

Electron microscopy (EM): Produce giga- and teravoxel images of biological structures at the nanoscale with SBF-SEM, FIB-SEM.

Light microscopy: Produce similarly-sized datasets of biological structures at the microscale with, e.g., cleared tissue light sheet microscopy.

Challenge: How do we scale microscopy research workflows for such datasets?



(**Top**) Human platelet sample imaged via SBF-SEM. LCIMB, NIBIB. (**Bottom**) Brain tissue sample imaged via cleared tissue light sheet microscopy. Section on High Resolution Optical Imaging, NIBIB.

ELECTRON MICROSCOPY IN LCIMB

The NIBIB's Laboratory of Cellular Imaging and Macromolecular Biophysics (LCIMB) develops EM techniques to investigate biological systems.

Newest hardware: serial block face scanning electron microscope (Gatan 3View).

Creates large (10^3 – $10^6\,\mu m^3$) image volumes by repeated cutting and scanning of a sample.

This talk: Walk through a SBF-SEM project workflow, highlight the challenges of scaling and automation.

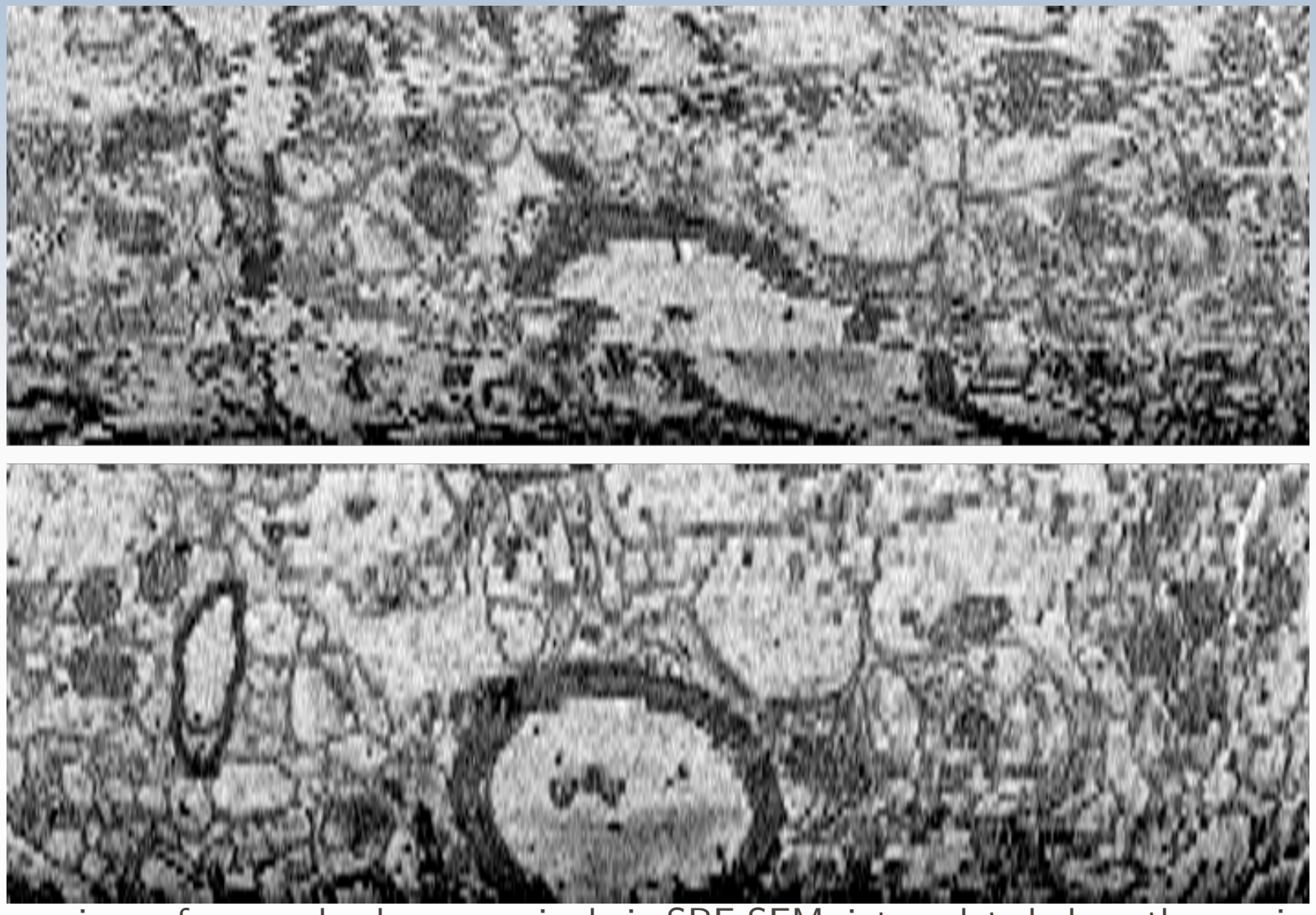
PREPROCESSING

Conversion: Data exits microscope in DM4 format. Convert to 32-bit TIFF using **Digital Micrograph**.

Background removal: Calculate background value, subtract from data using **ImageJ**.

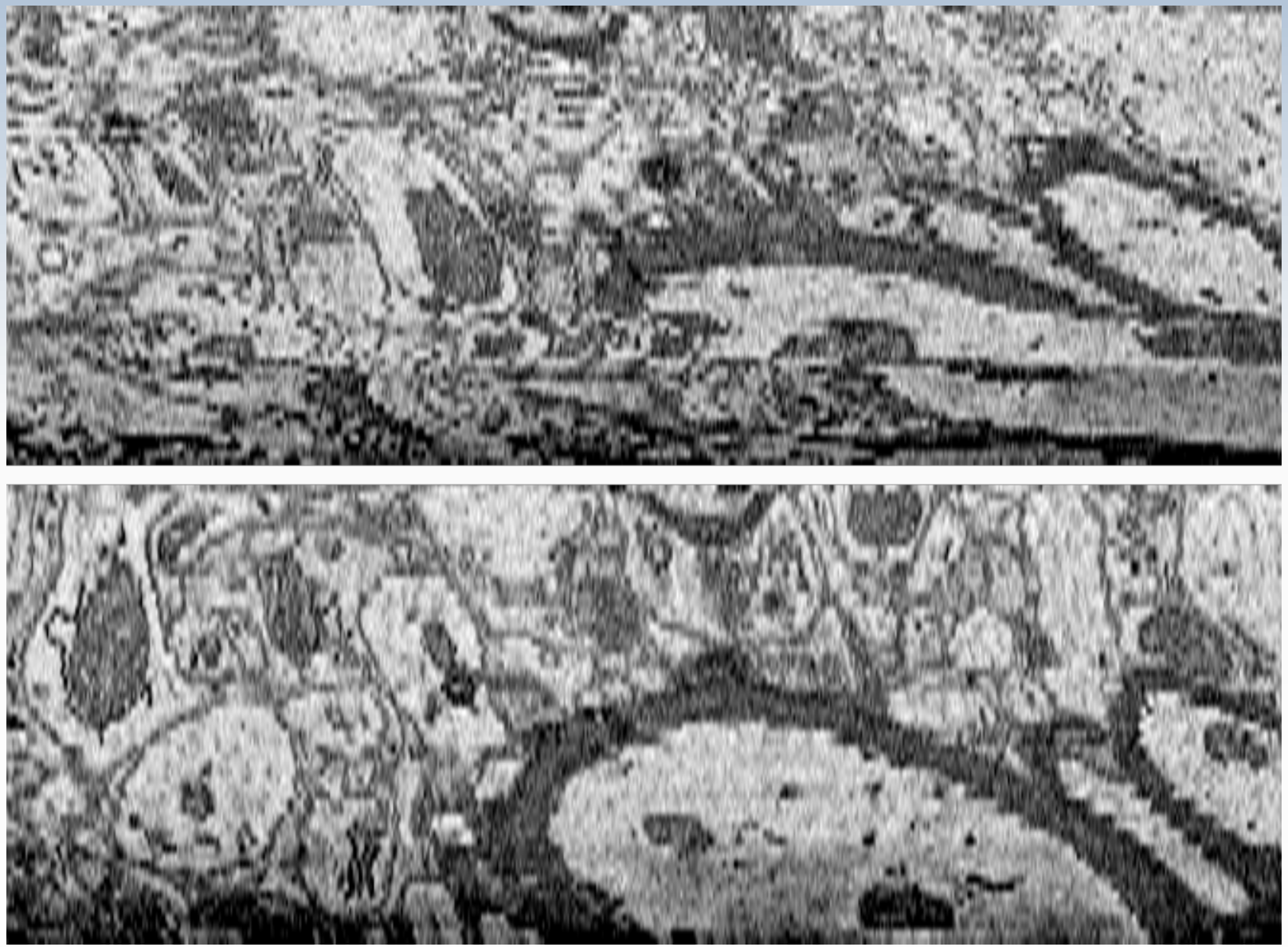
Alignment: SBF-SEM slices are misaligned along the slice axis. Perform multiple linear stack alignments - coarsely at large scales, finer for smaller ROIs - using **IMOD**.

ALIGNMENT EXAMPLE: $X - Z$



$x - z$ views of a neural volume acquired via SBF-SEM, interpolated along the z axis to approximate isotropy. (**Top**) Unaligned volume view. (**Bottom**) Aligned volume view.

ALIGNMENT EXAMPLE: $Y - Z$



$y - z$ views of a neural volume acquired via SBF-SEM, interpolated along the z axis to approximate isotropy. (**Top**) Unaligned volume view. (**Bottom**) Aligned volume view.

PROCESSING

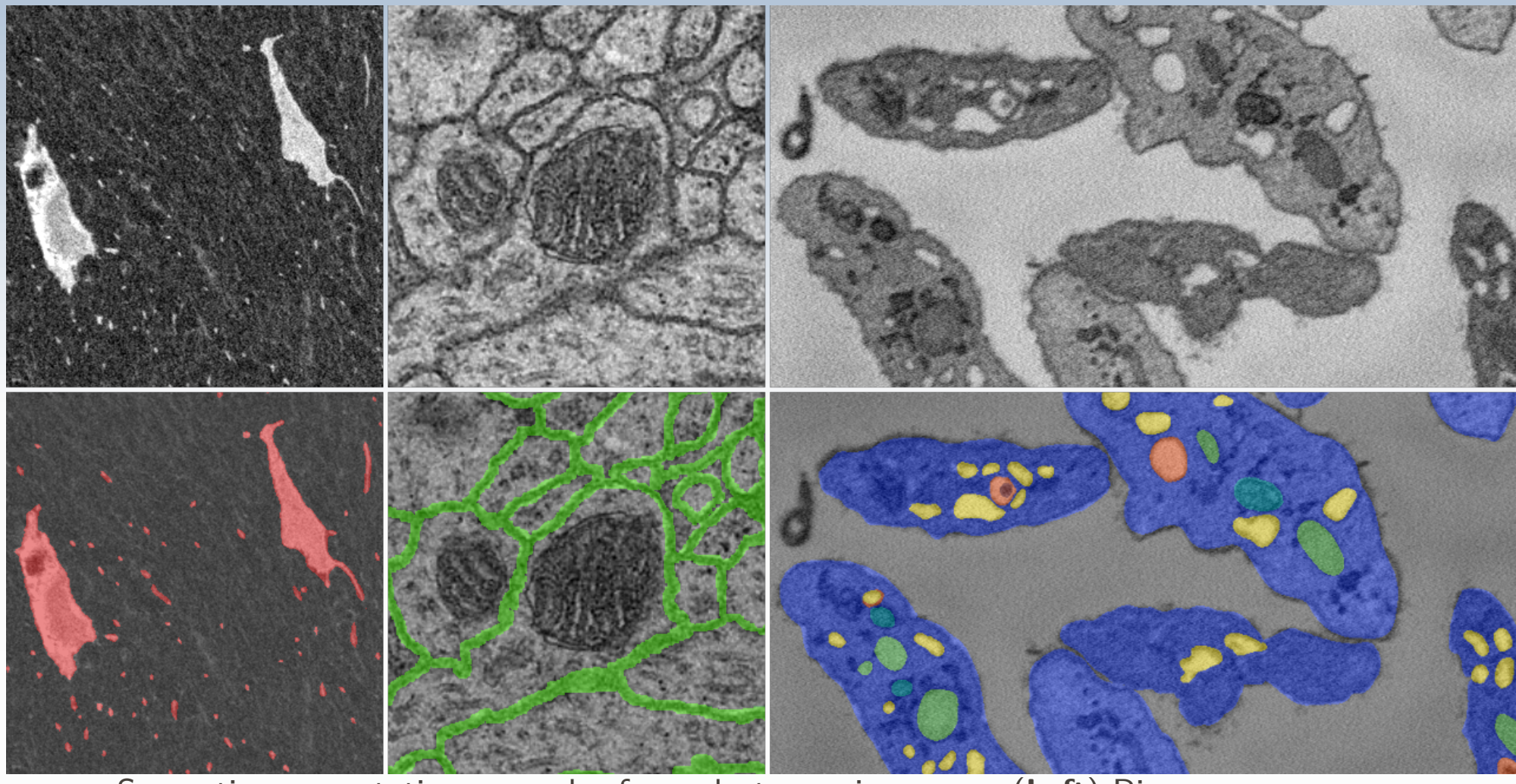
Semantic segmentation: Assign a class label to each voxel in an image (background, cytoplasm, mitochondria, etc.).

Instance segmentation: Assign a unique label to each object in an image.

Combine the two to convert raw image pixels to lists of objects, the regions they encompass, and their semantics. (i.e. vision??)

Currently done by hand in **Amira**. Very tedious, impossible to apply to full SBF-SEM datasets.

SEGMENTATION EXAMPLES



Semantic segmentation examples from electron microscopy. (**Left**) Binary segmentation of osteocyte soma and projections. LCIMB. (**Center**) Binary segmentation of neuron membranes. ISBI 2012 segmentation challenge. (**Right**) 7-class segmentation of platelet cells and organelles. LCIMB.

ANALYSIS

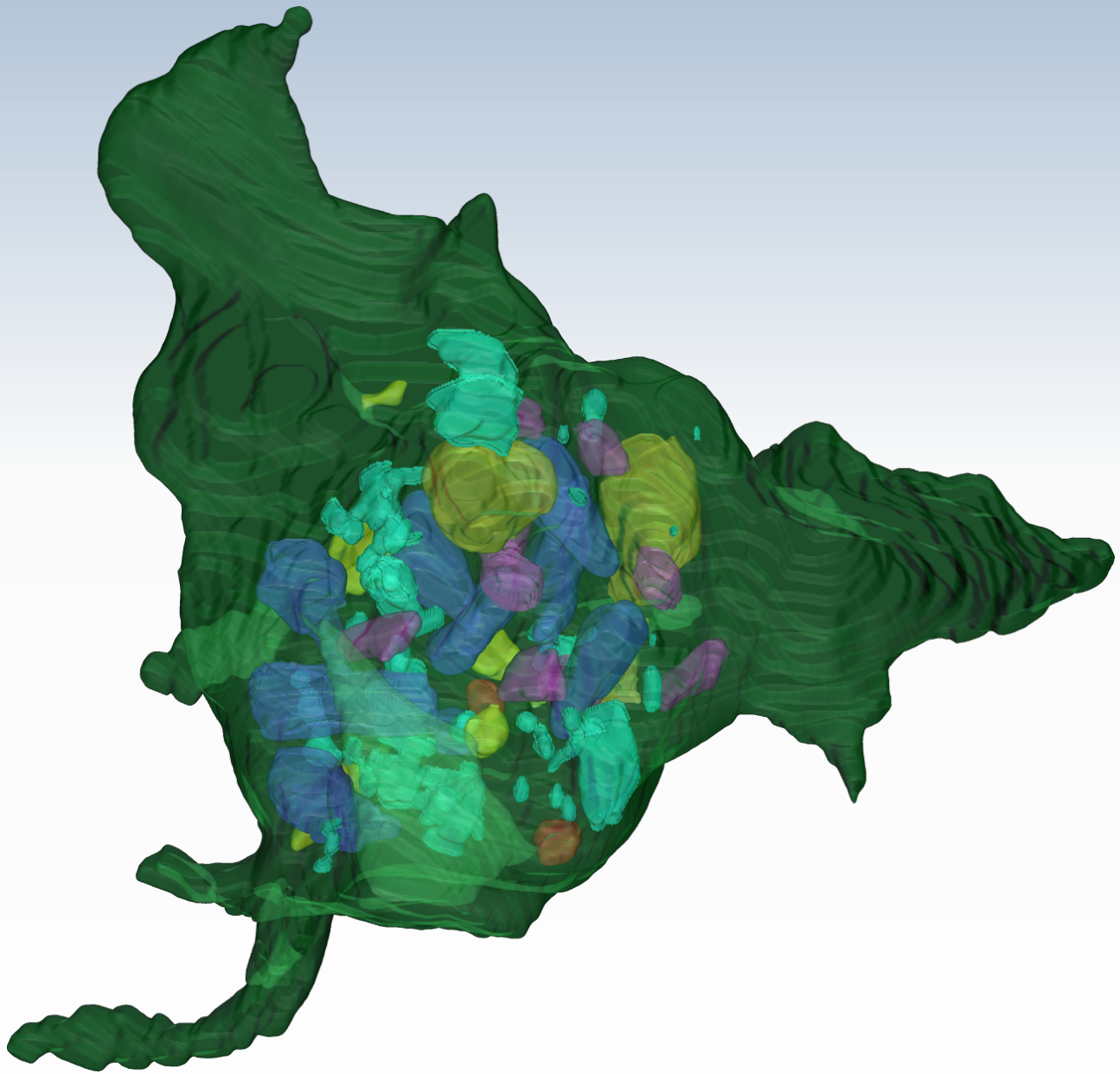
Goal: Derive quantitative information about the structure and arrangement of cells and organelles.

Segmentation information can be used to, e.g.:

- Compute volumes and surface areas of cells and organelles.
- Compute spatial distribution of cells within a tissue, or organelles within a cell.
- Other morphological analysis (do activated platelets look different from inactivated platelets?)
- Create 3D visualizations of image data.

Biologists tell me this is useful for biology.

VISUALIZATION EXAMPLE



3D rendering of the semantic segmentation of a platelet cell.

AUTOMATION CHALLENGES

Preprocessing and processing need to be automated to scale up to full SBF-SEM datasets. (Analysis + visualization are easier).

Little challenge: automating alignment. Rigid multiscale alignment works best for small ROIs.

LCIMB is investigating elastic alignment methods to improve this.

AUTOMATING SEGMENTATION

Big challenge: automating segmentation with computer vision tools.

Semantic and instance segmentation automation difficulties are problem-dependent.

Significant progress in recent years using deep learning (DL), but research problems remain.

AUTOMATING SEMANTIC SEGMENTATION

Simple segmentation problems (high contrast, low-noise) can be solved with thresholding

Most EM problems of interest are not so simple.

Relevant DL modules for difficult semantic segmentation:

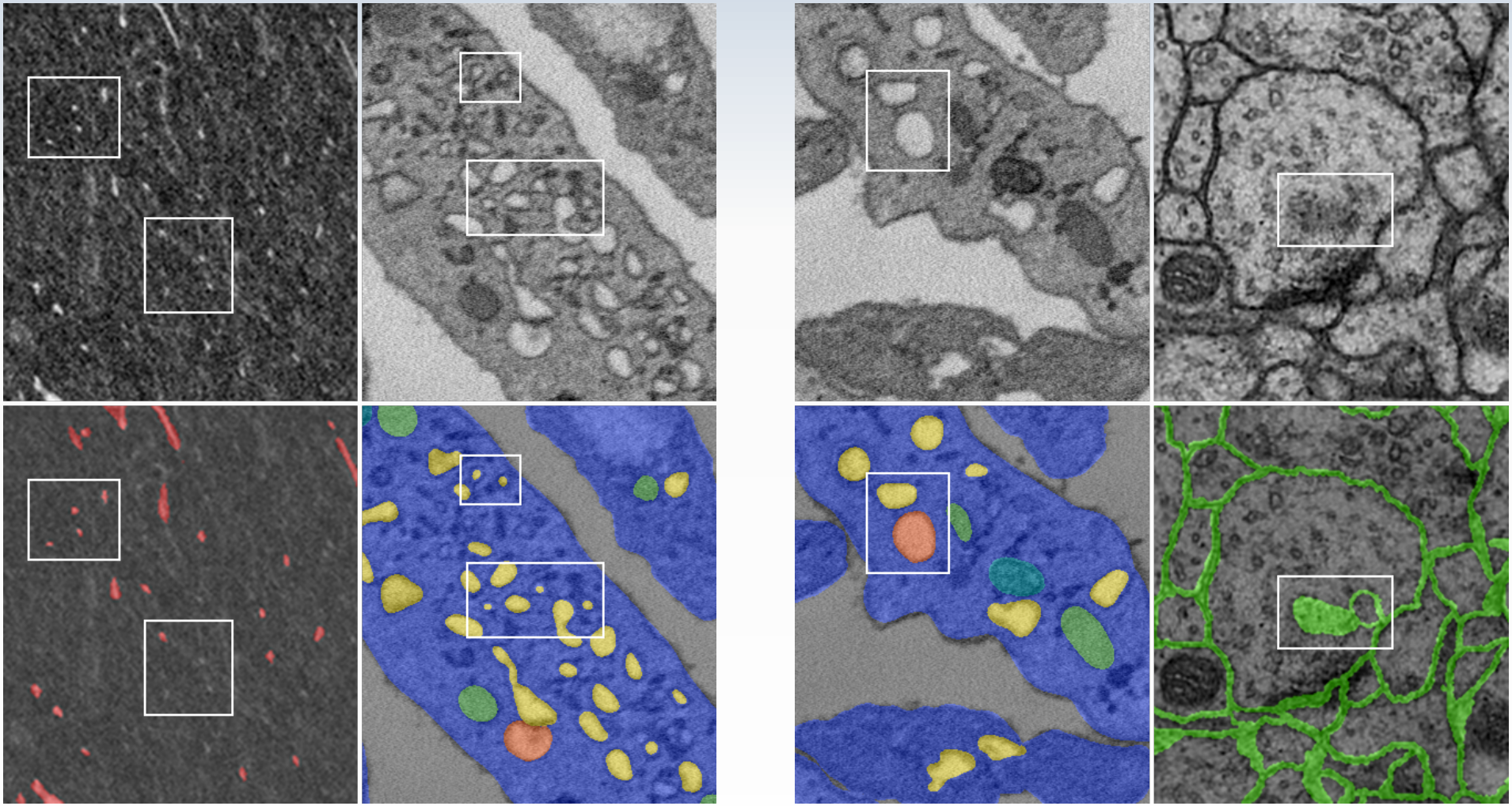
- Encoder-decoder networks (U-nets)
- Dilated convolution networks (context aggregation nets)
- Spatial pyramid pooling modules (ASPP).

SEMANTIC SEGMENTATION CHALLENGES

A practical semantic segmentation algorithm requires high ($> 99.9\%$) accuracy despite:

Noise + small objects

Difficult label assignment



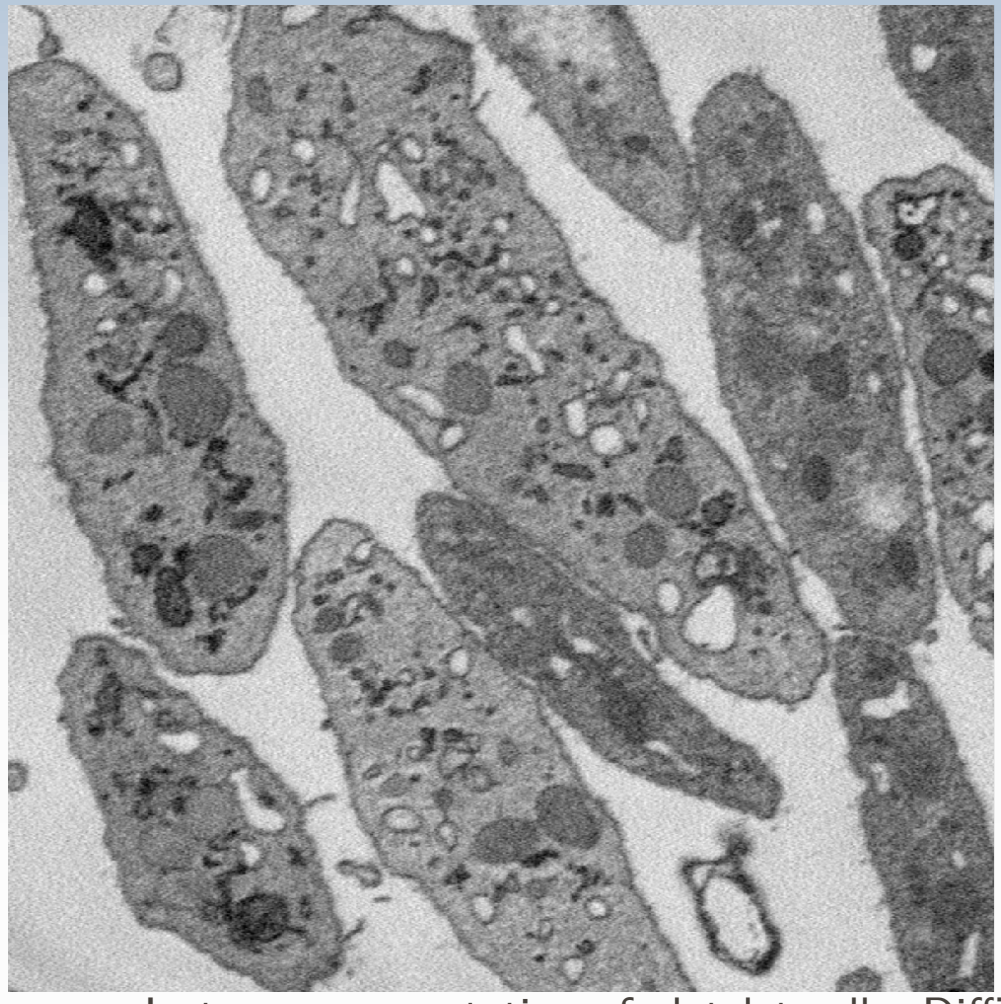
AUTOMATING INSTANCE SEGMENTATION

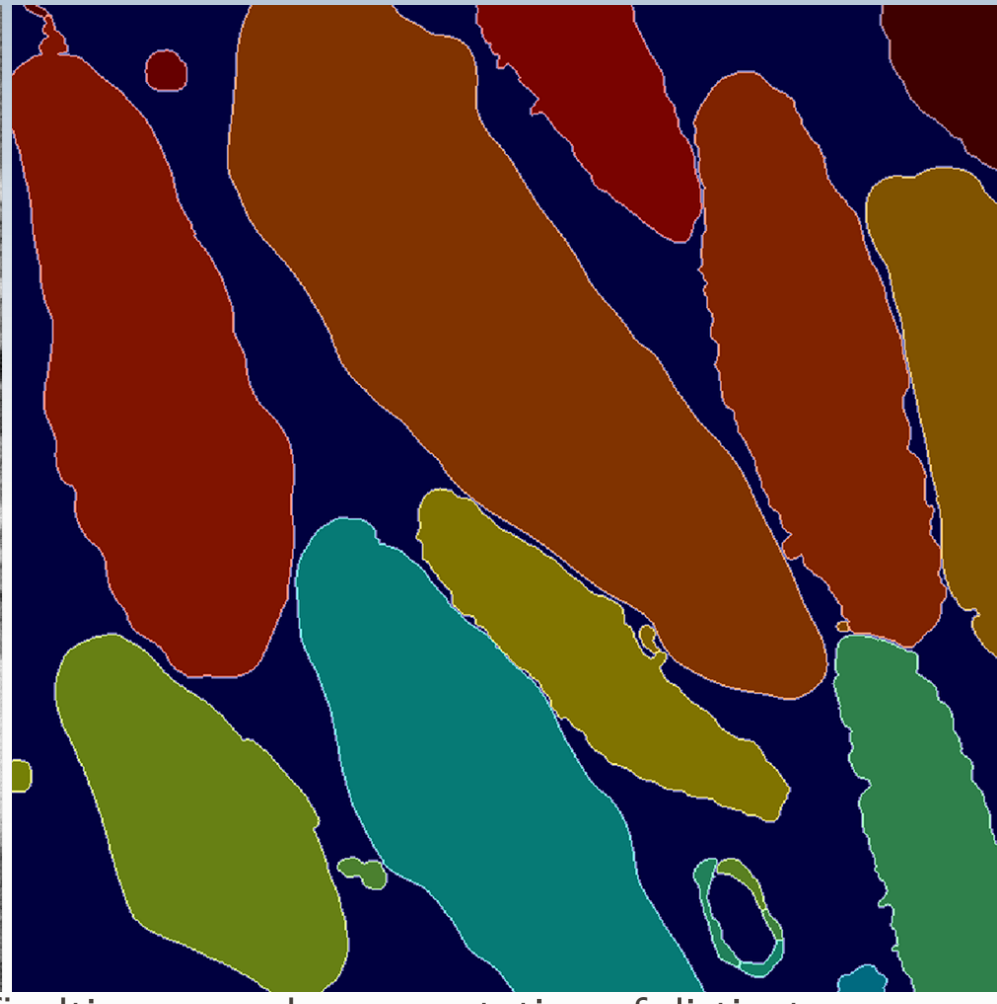
Main challenge, given semantic segmentation info: separating objects of the same "type", despite contact.

Distance-transform watershedding works for spherical objects, what about everything else?

I am less knowledgeable on DL state of the art here, but check out Mask R-CNN as a starting point.

INSTANCE SEGMENTATION CHALLENGES





Instance segmentation of platelet cells. Difficulties are undersegmentation of distinct objects in contact, as well as oversegmentation of objects with "pinch points".

SCALING CHALLENGES

Separate but related: How to scale an automated processing pipeline?

Large-scale image processing is well-suited to high-performance computing (HPC) systems like Biowulf, but adapting workflows to HPC can be challenging.

How to move workflow components from desktop software (IMOD, ImageJ) to scripting software and compiled languages suited to HPC (Python, C++)?

- Techniques that work for small datasets may not scale efficiently to large ones (computational complexity).

How to manage of large datasets? In-lab, transfer to and from HPC systems, sharing with collaborators.

NEURAL NETWORK SCALING CHALLENGES

Training neural networks on large datasets may require distributed training across multiple compute nodes.

Architecture selection: Choosing one or more neural net architectures for a project. Expensive to evaluate, again requiring distributed training.

Large, multi-module architecture possibilities grow combinatorially with the number of potential modules.

SCALING SOLUTIONS

Lots of scaling problems, but what are the solutions?

NIH HPC facility provides tutorials, references, and classes to make it easier to get started on Biowulf.

Data management, workflow automation, and neural network training and deployment challenges remain.

SCALING SOLUTIONS

NVIDIA has recently begun partnering with NIH labs to help solve workflow automation and acceleration challenges.

Question: Can we work together to better harness big data for biomedical microscopy?

Many NIH microscopy labs working independently on similar problems. Let's solve our shared challenges together, with the help of computational software and hardware experts!

Are you interested? Please take this survey to help us figure out what can be done: https://goo.gl/forms/5Fi8bZfBUb5hDOK13