

# High Throughput Connectomics

Nir Shavit  
*MIT, Tel-Aviv University,  
and Neural Magic Inc*

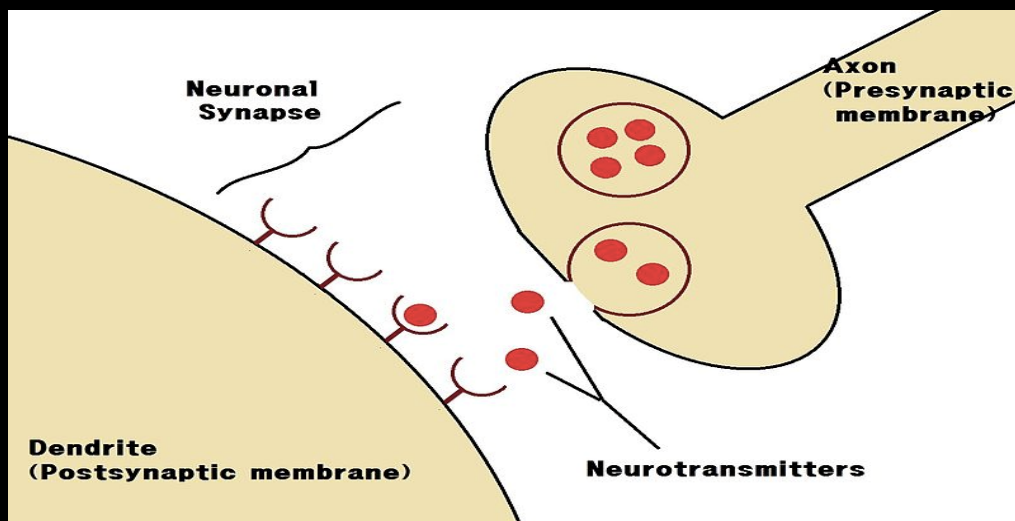
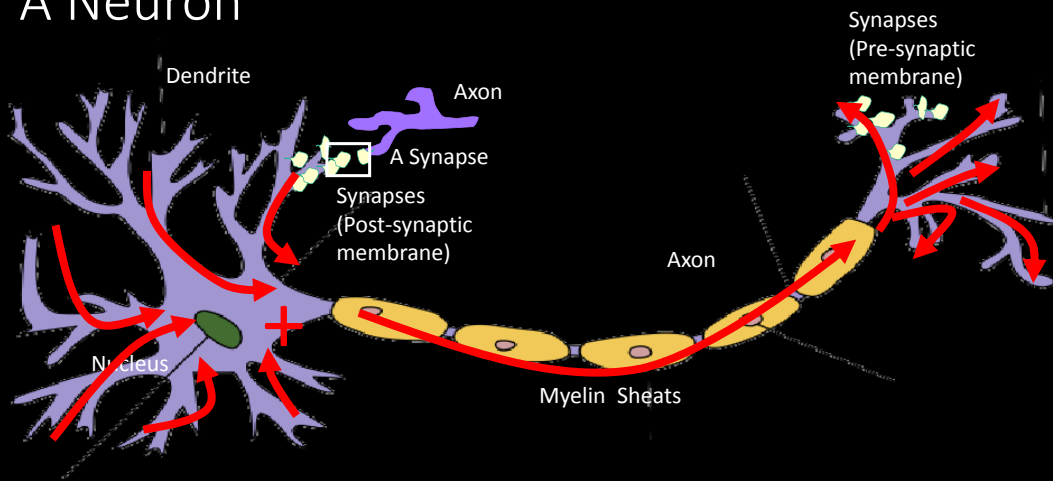
## The Cray 1 SuperComputer



1/10 of a second

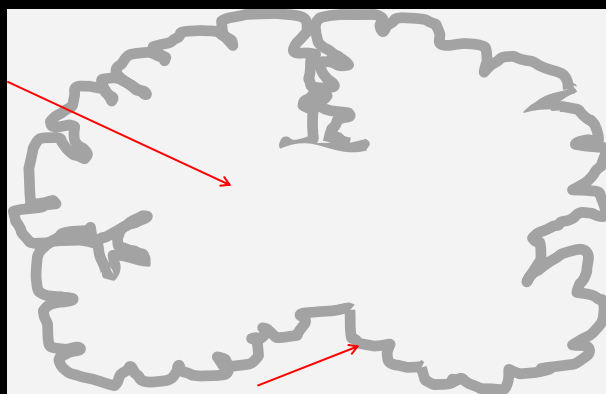


## A Neuron



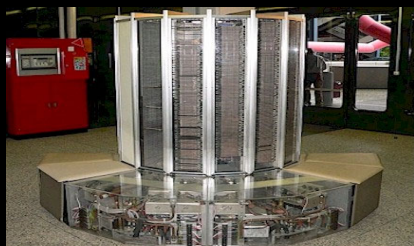
## Brain: White and Grey

White Matter



Grey Matter

## White and Grey



Grey



White

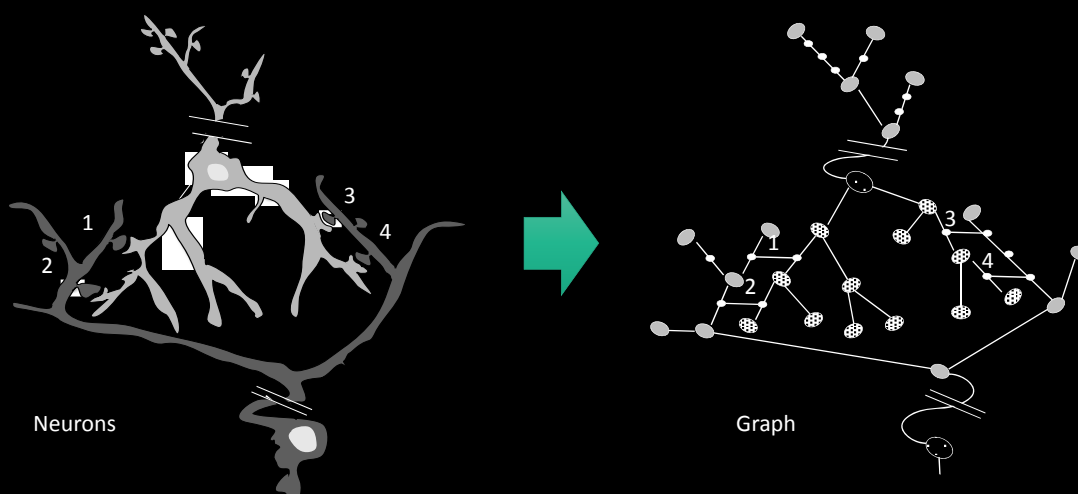
1/10 of a second



About 20 neurons fire in sequence

Network of neurons has “ridiculous”  
levels of parallelism

My Goal: Understanding this parallelism



# Connectome = Brain Graph

Fly: 100 thousand neurons (nodes) x  
100 synapses (per node) =  
10 Million edges

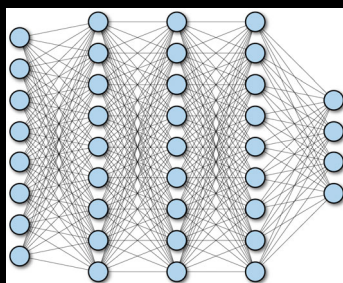
Mouse Cortex: 4 million neurons (nodes) x  
10 thousand synapses (per node) =  
40 Billion edges

Human Cortex: 30 billion neurons (nodes) x  
10 thousand synapses (per node) =  
300 Trillion edges

## Implications for Machine Learning

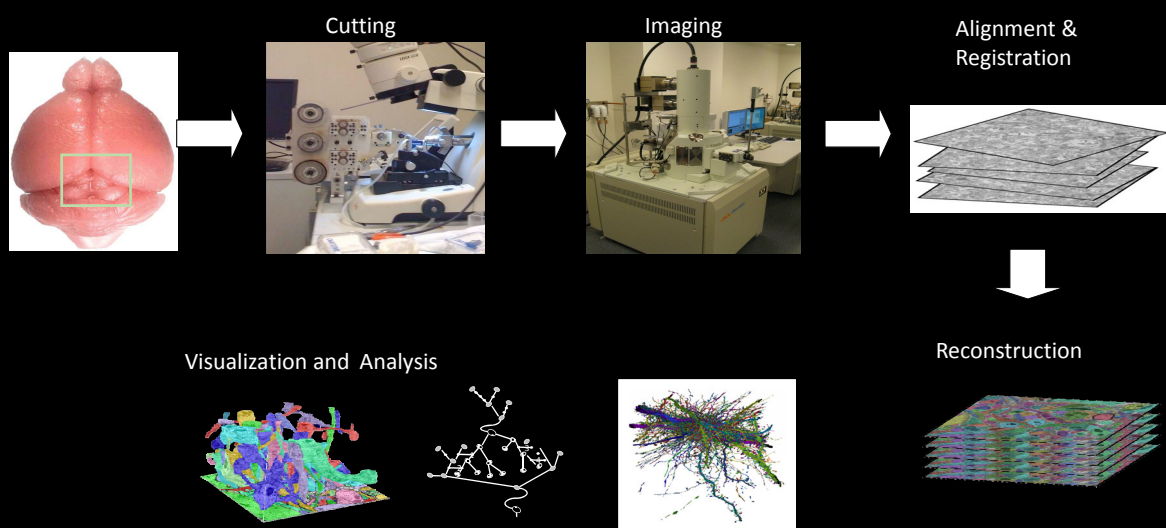
Common wisdom: Deep learning is compute intensive like brain –  
need special accelerator hardware like GPUs or TPUs ...

A Fully connected  
neural network

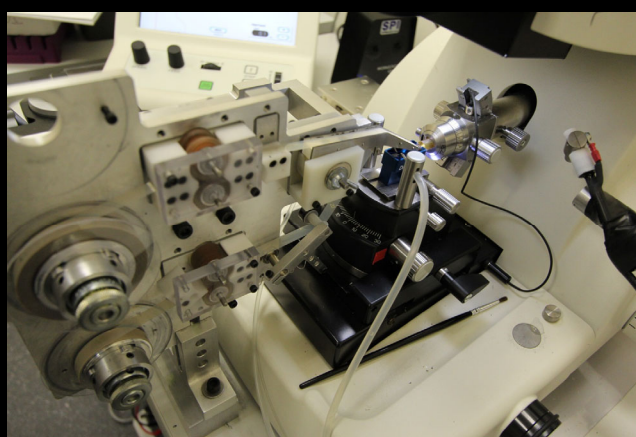


Our cortex is very sparse: ...understanding its structure will be key to  
efficient ML as we move forward to mimicking brains...

## Mapping a Connectome

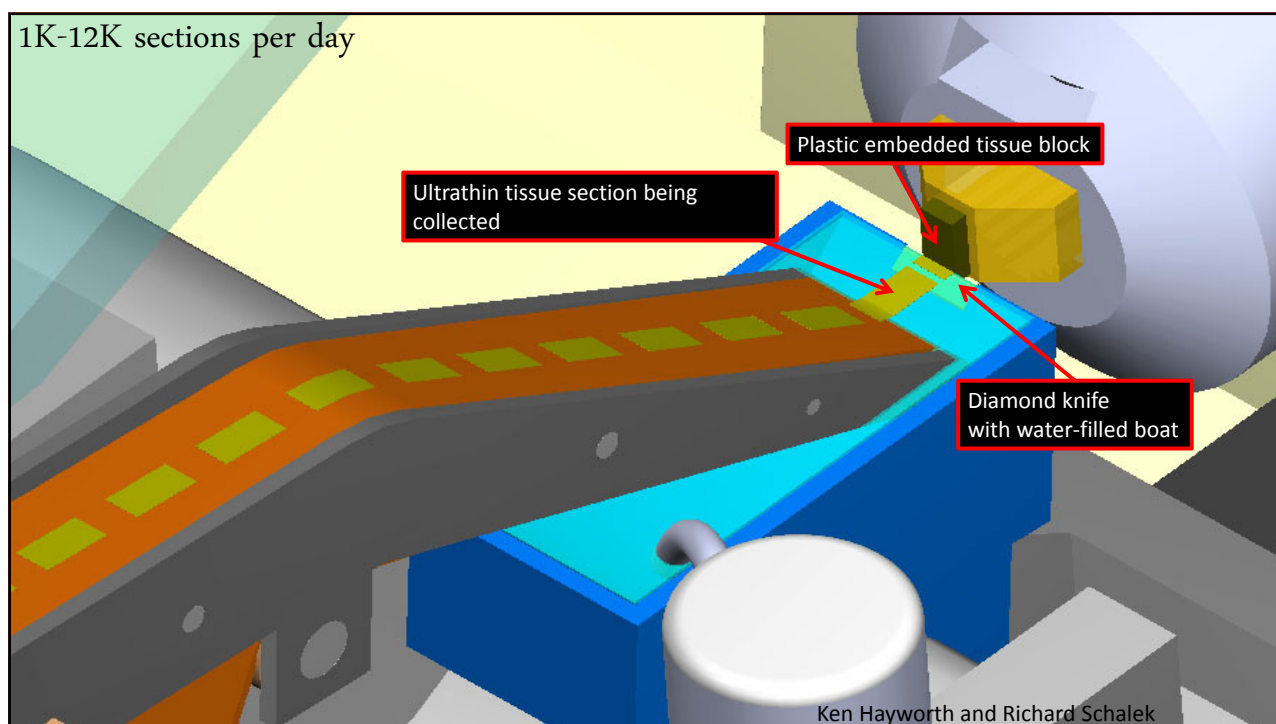


## Automatic Tape-collection Mechanism for Ultramicrotomes



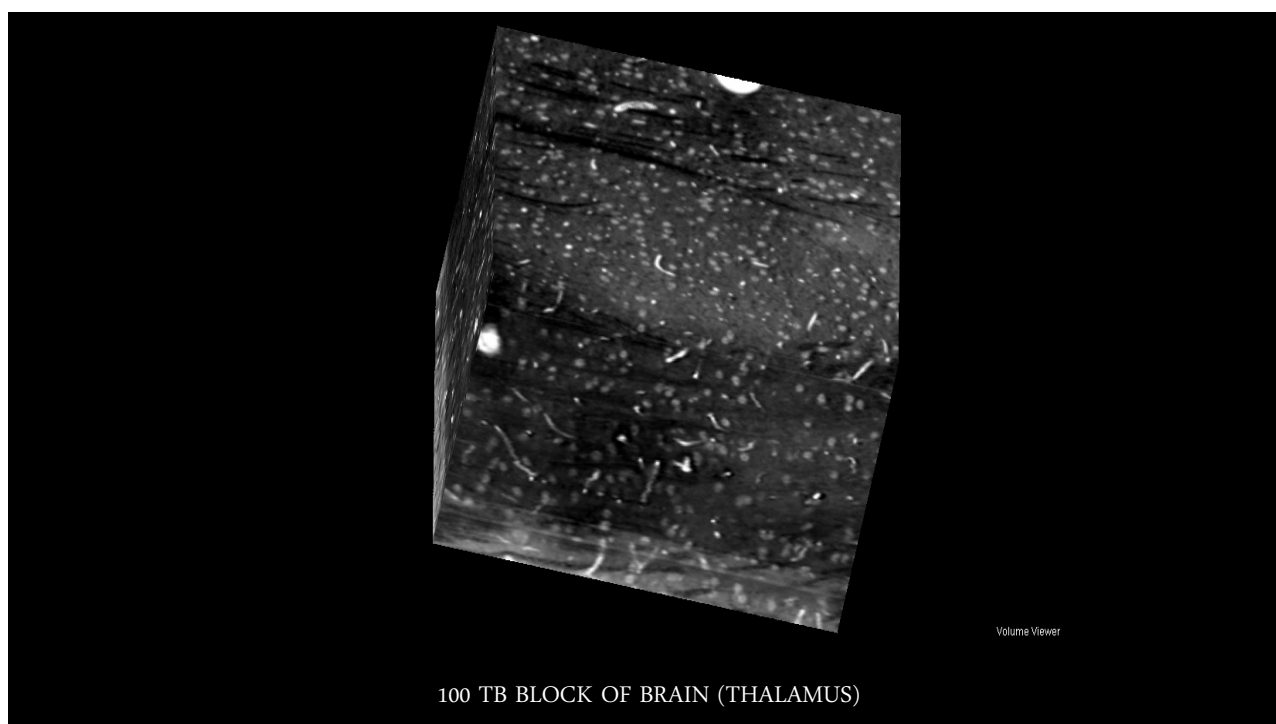
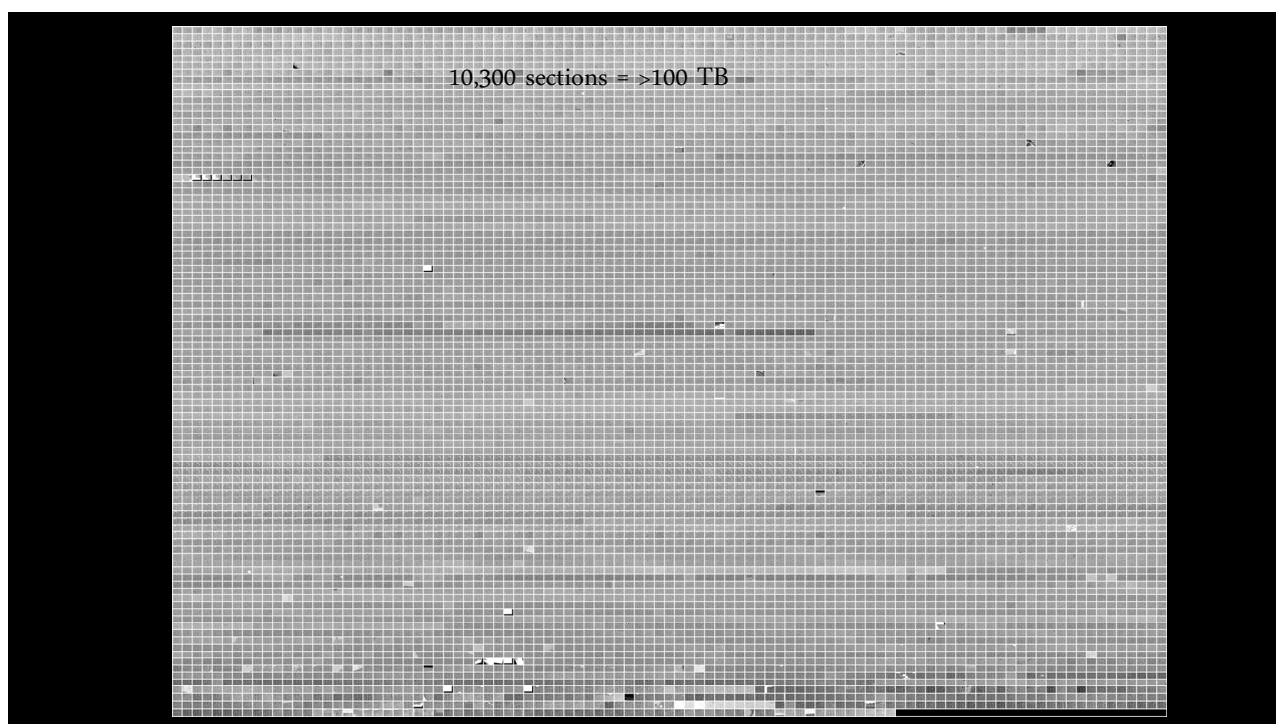
Lossless, Nondestructive, Sturdy sections

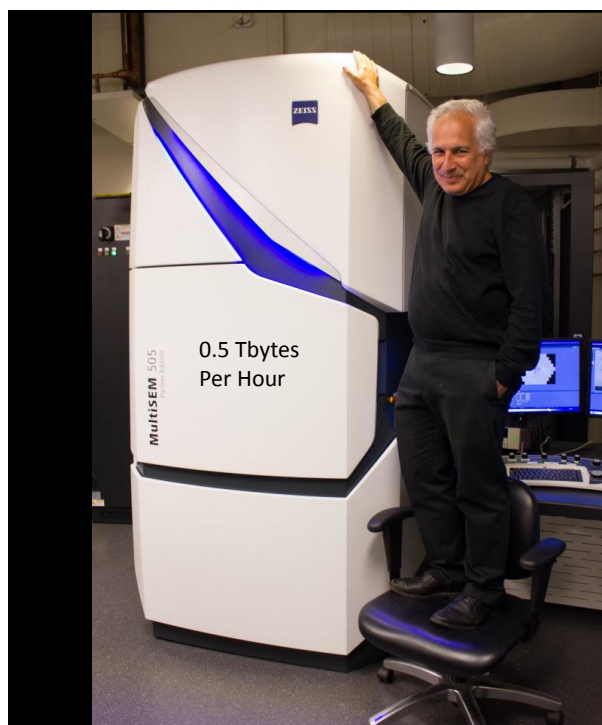
Ken Hayworth and Richard Schalek



## Brain on tape







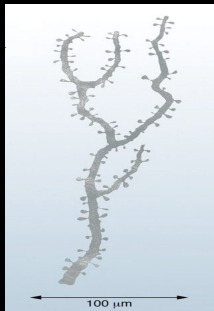
**Past:** Full EM based connectome of nematode *C. Elegans* 302 Neurons 7500 synapses in 10 years (Brenner et al in 1986)

**Today:** single multibeam microscope will image 1 cubic mm in a matter of months

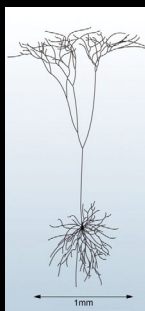
**The Bottleneck:**  
Reconstruction of the connectome from the EM images

## “Really Big” Data

10k  $\mu\text{m}^3$   
10 TByte



1 mm<sup>3</sup>  
2 PByte



Walmart Database  
~2.5 Petabytes

Our current target

500 mm<sup>3</sup>  
1 EByte



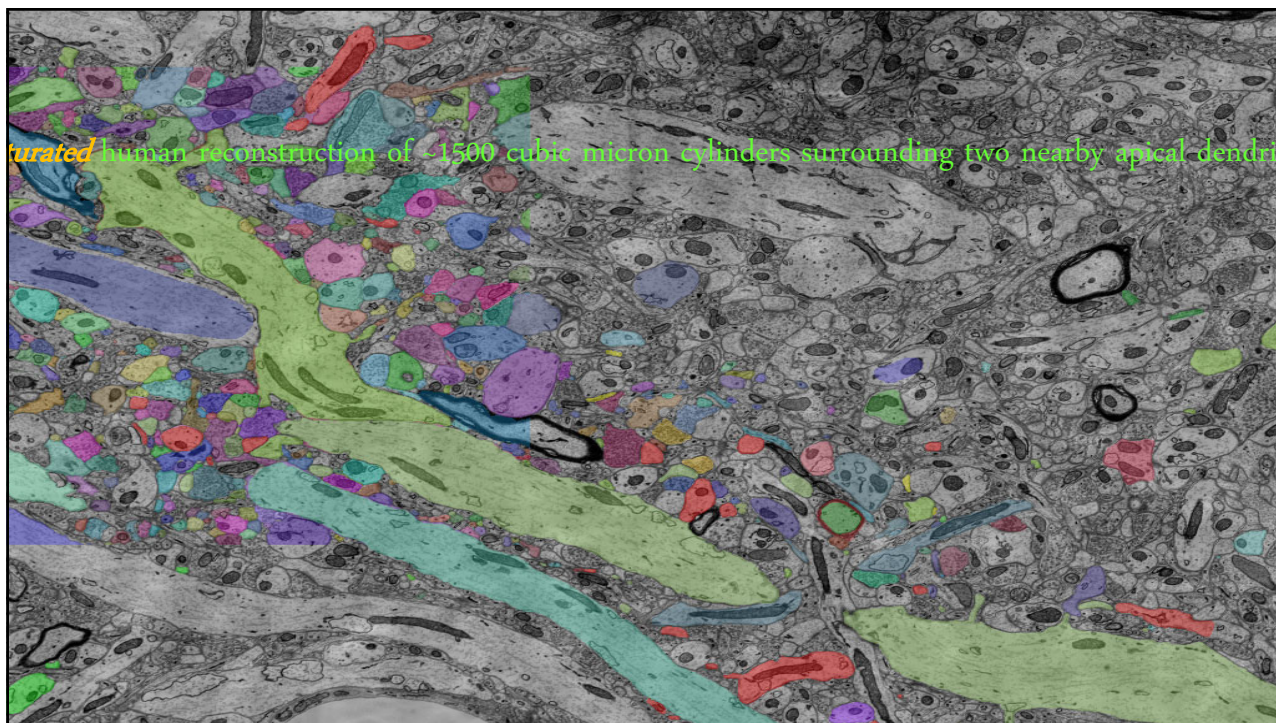
Rodent

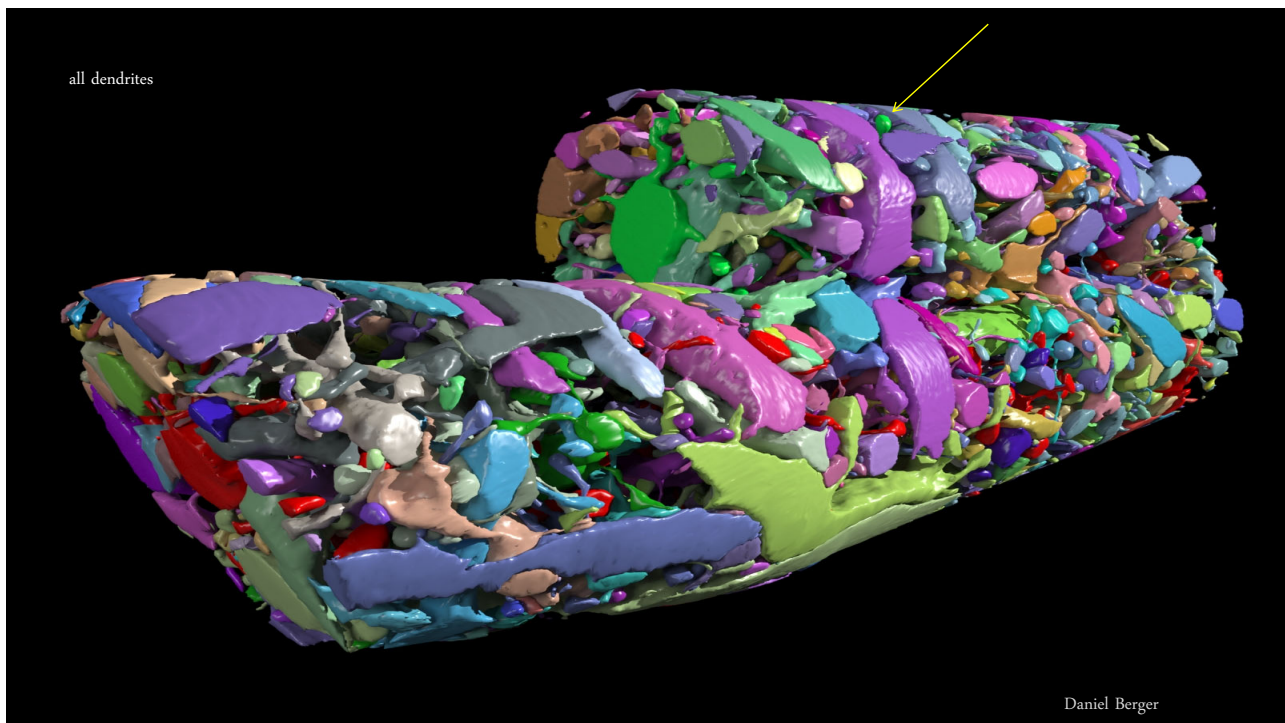
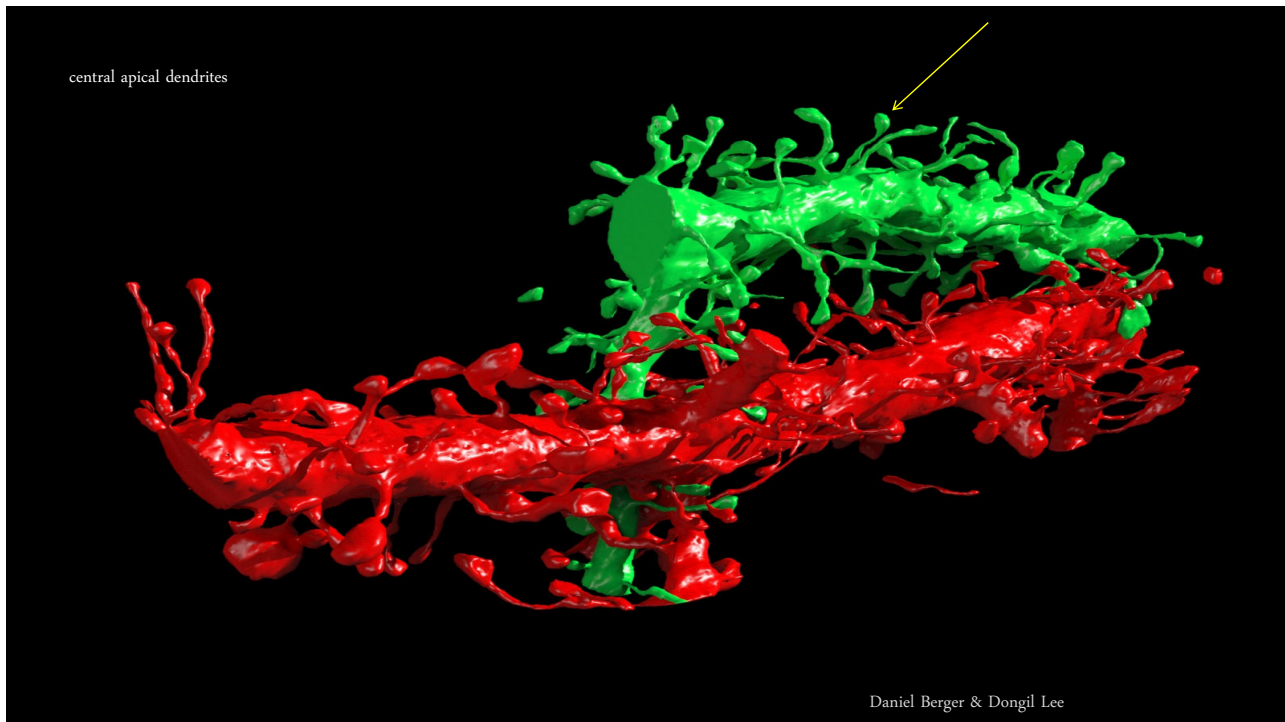
Our next target

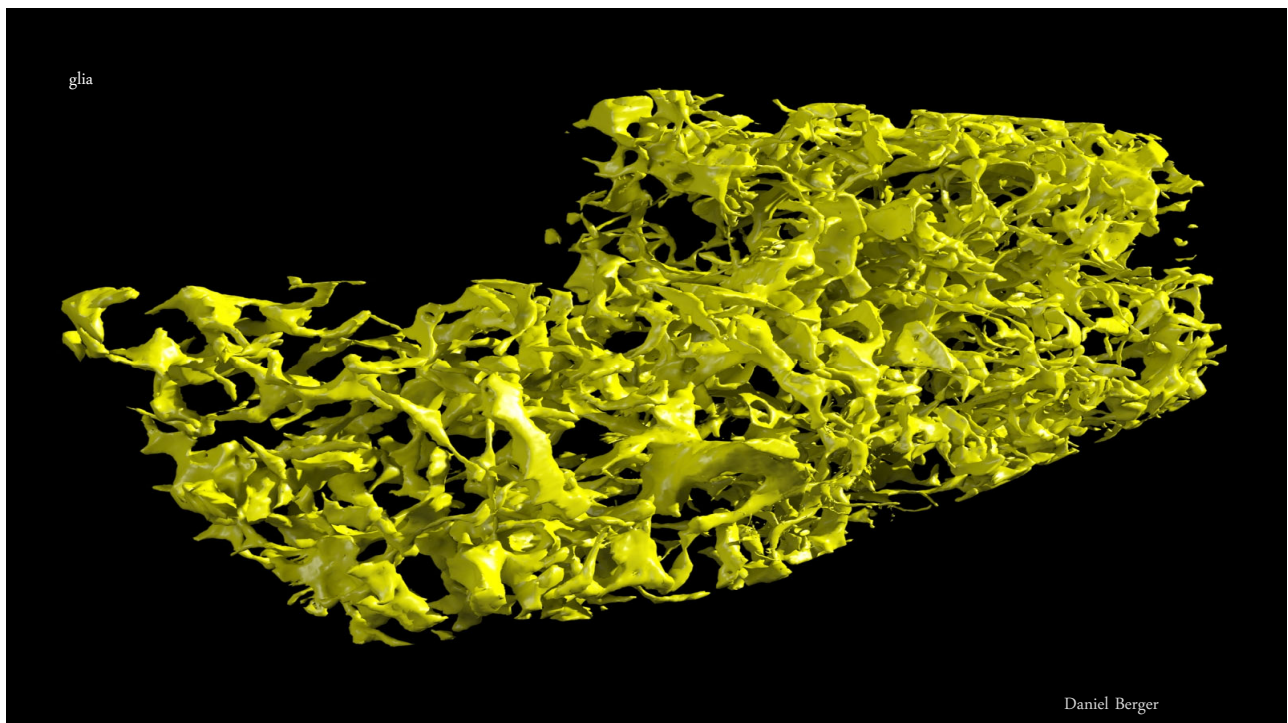
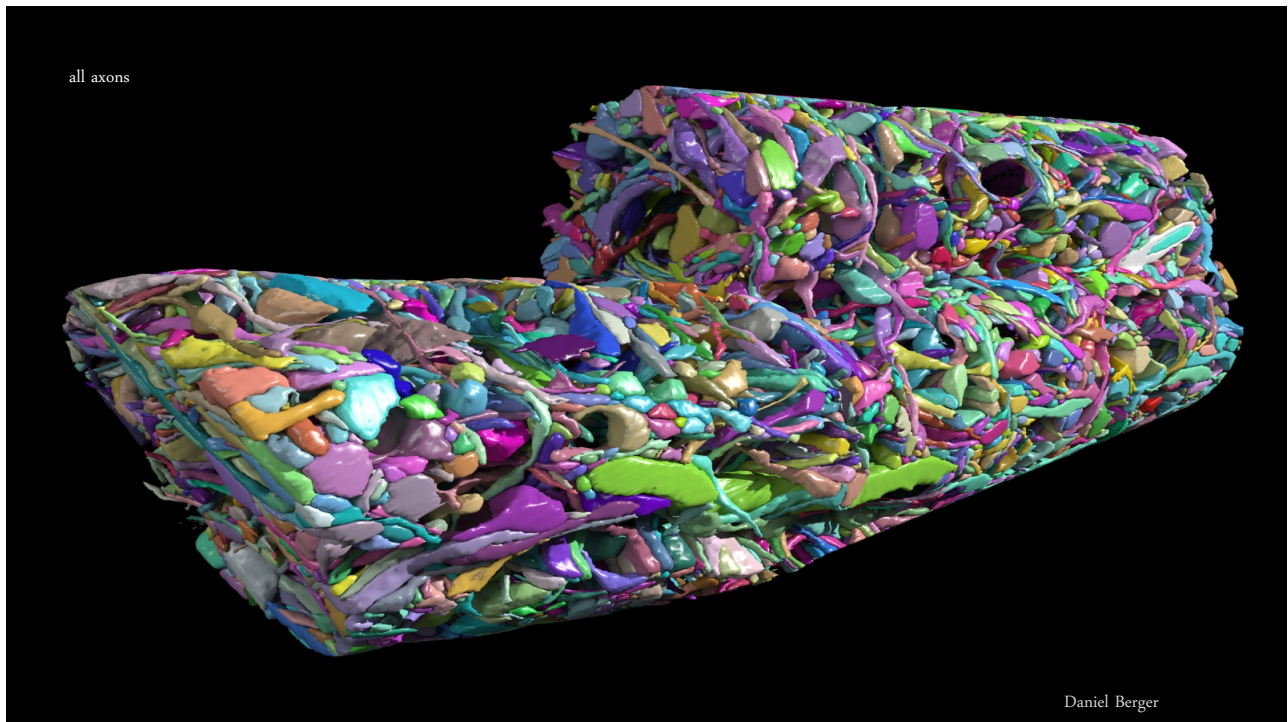
1.4 M mm<sup>3</sup>  
1 Zbyte

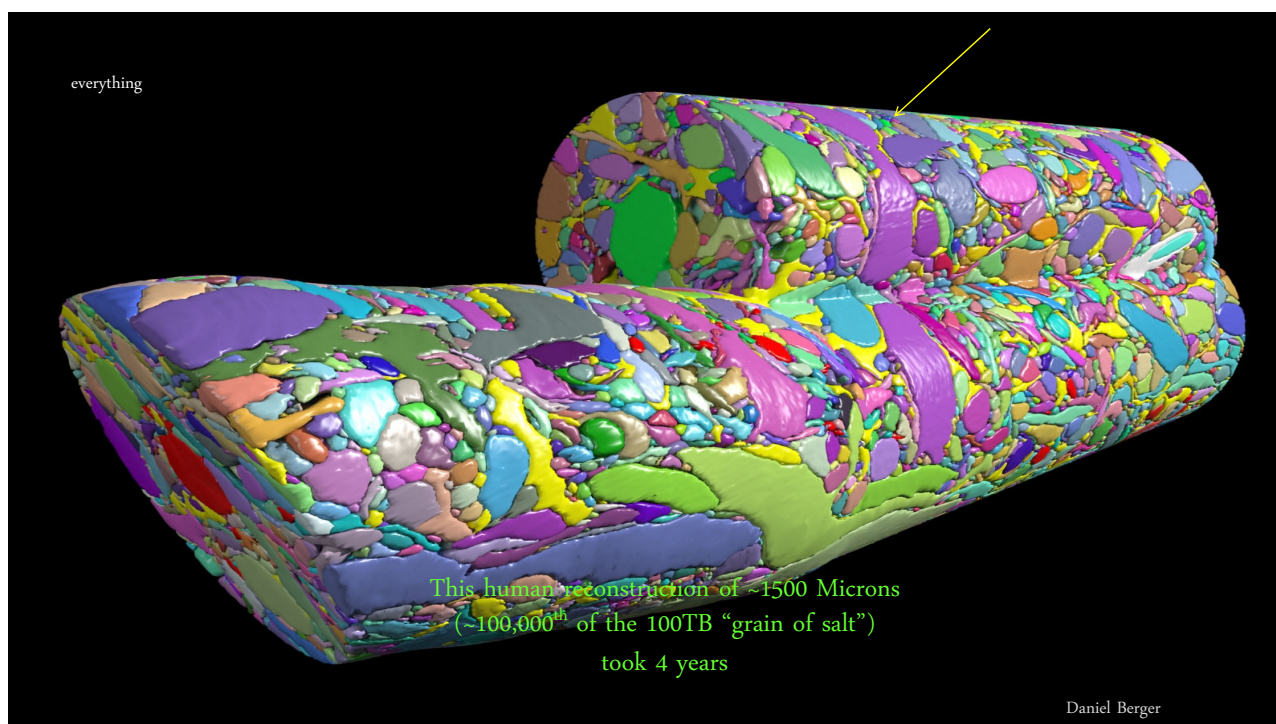
Human

1 Zbyte all info  
stored in world

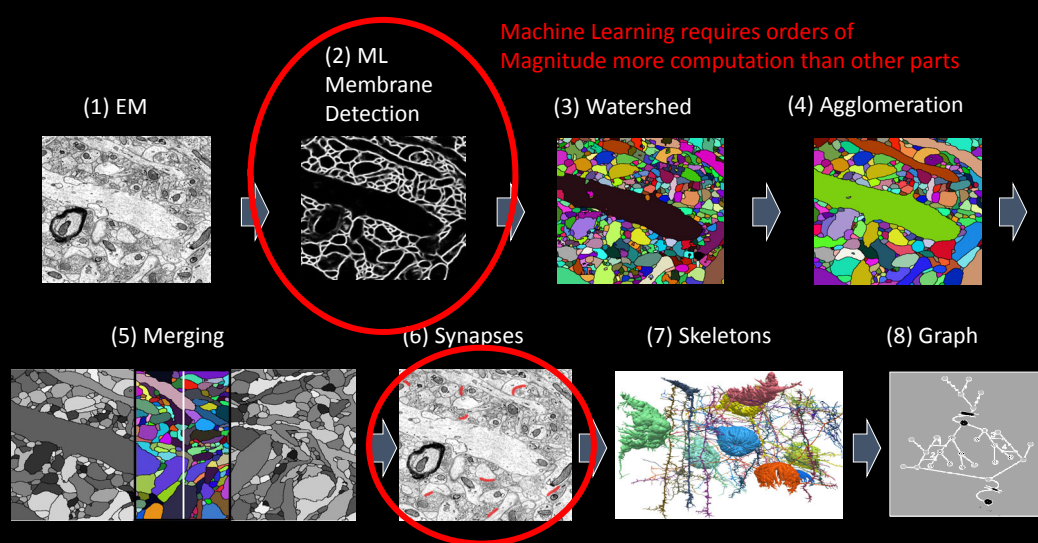


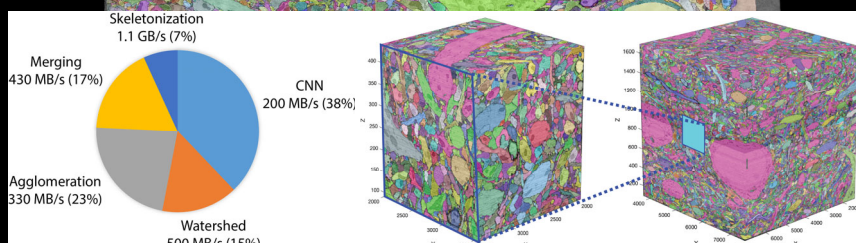






## "Traditional" Automated Reconstruction Pipeline





Yaron Meirovitch

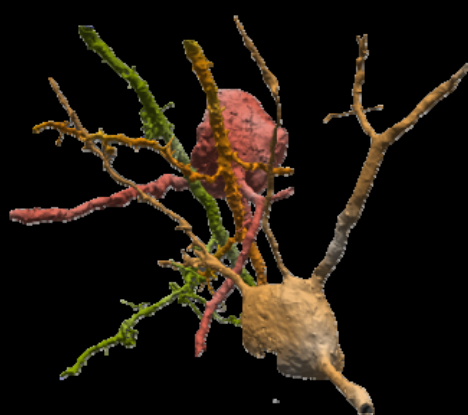


Alex Matveev

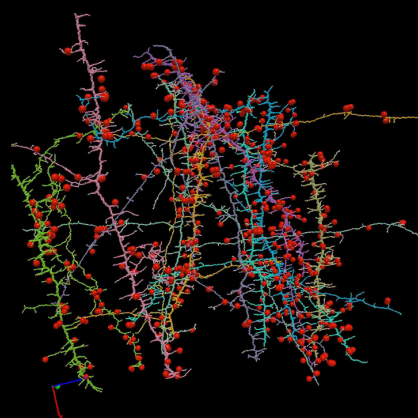
Pipeline run on .473 TByte Kasthuri et al. Dataset.

On standard 72 core machine  
~4 hours/TB (only 2x slower  
than Multi-Beam Microscope)

# Traditional Agglomerating Pipeline: Great reconstructions but still too many errors



Matveev et al



Santurkar et al

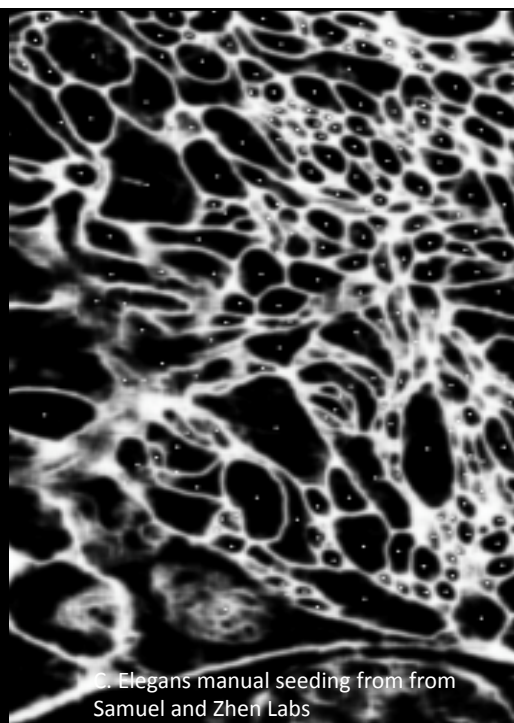
## Try to do what neurobiologists do...

- First: Skeletonization
- Then: Expansion of objects

(Unlike existing automated pipelines that create expanded objects and then agglomerate them)

## Skeletonization

- Human:
  - Neurobiologist skeletonizes objects perfectly
  - By picking one seed per object per slice
- Automatic:
  - To Date: Primarily by shrinking a dense segmentation
  - New Approach: learn to pick one seed per object based on EM and skeletonization of prior slices

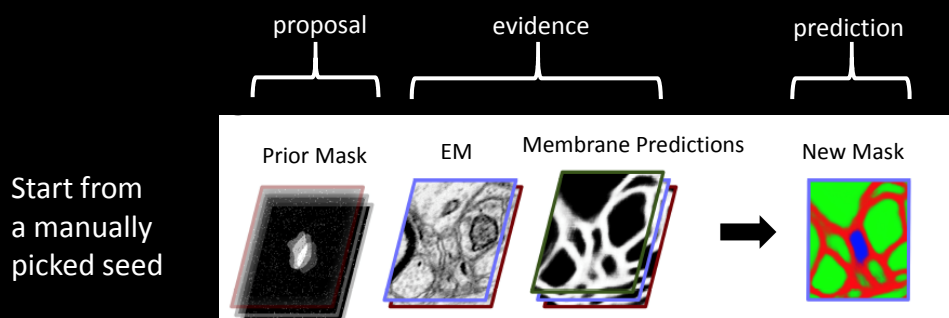


C. Elegans manual seeding from from Samuel and Zhen Labs

## Mask Extend: Mimic Human Tracing of a Single Neurite



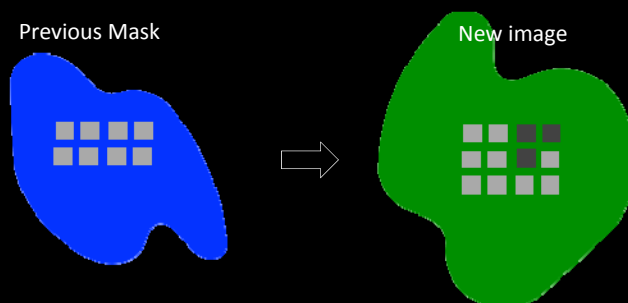
Yaron  
Meirovitch



Use CNN on XNN, our proprietary multicore CPU based ML execution engine.

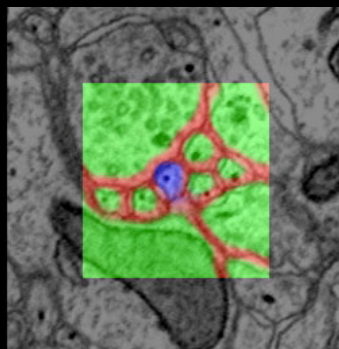
## Mask Extend: Mimic Human Tracing of a Single Neurite

Classification: For each pixel, does it belong to the partially reconstructed object?



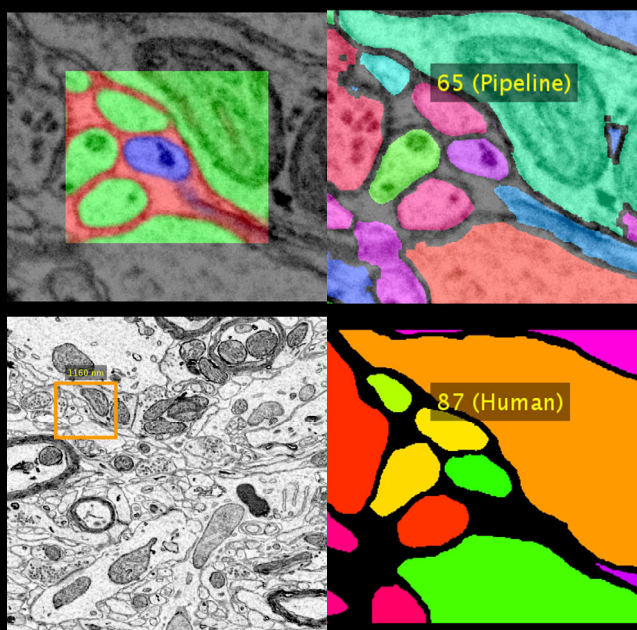
\*Independently (similar idea) by Januszewski et al. They use a much more powerful flood-fill network and no membrane predictions.

# Mask Extend



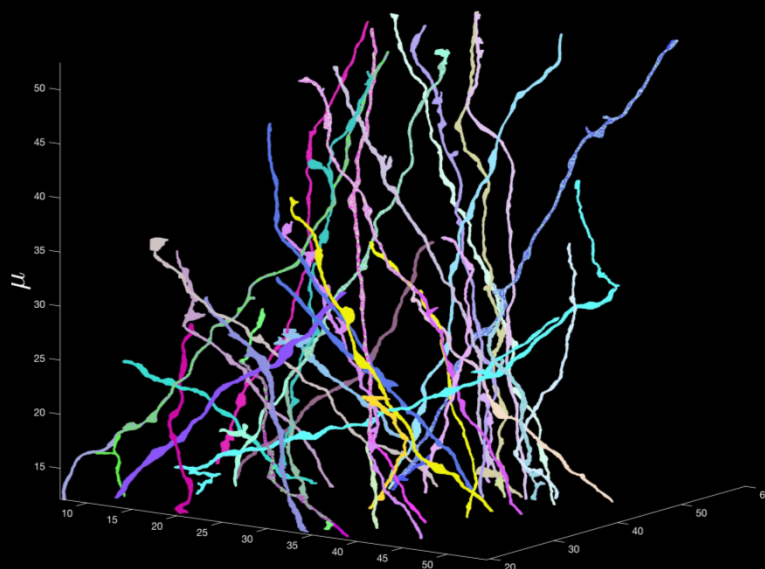
## Mask Extend

Corrects UNET  
CNN results on RAT  
ECS dataset from  
state-of-the-art  
“traditional”  
Pipeline

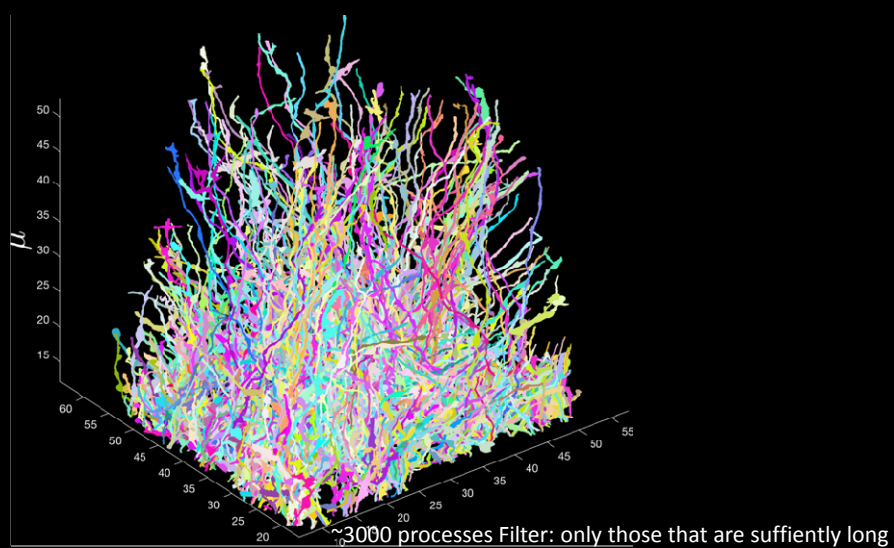


\*Volume from 405 upward is estimated ~77,500 cubic microns

## Single Neuron MaskExtend Classification



## Whole volume by repeated Mask Extends....



## MultiMask [ICML'19]

Imagine we can learn how to segment all neurons in next slice based on the partial segmentation of prior slice.



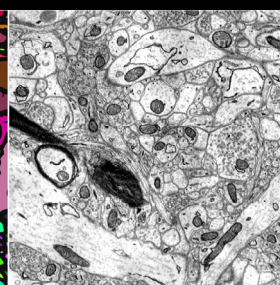
Yaron Meirovitch



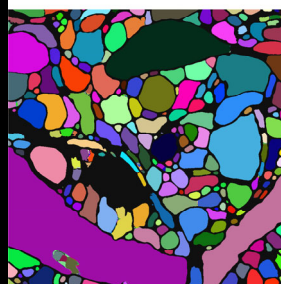
Lu Mi

Input  
(Seeds+EM)

Output  
(Segmentation)

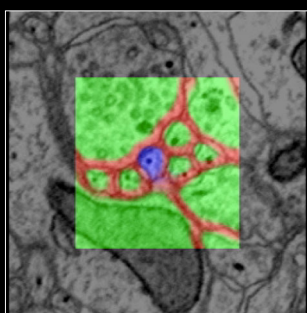


Prior



Next

## Classifying Unknown Classes

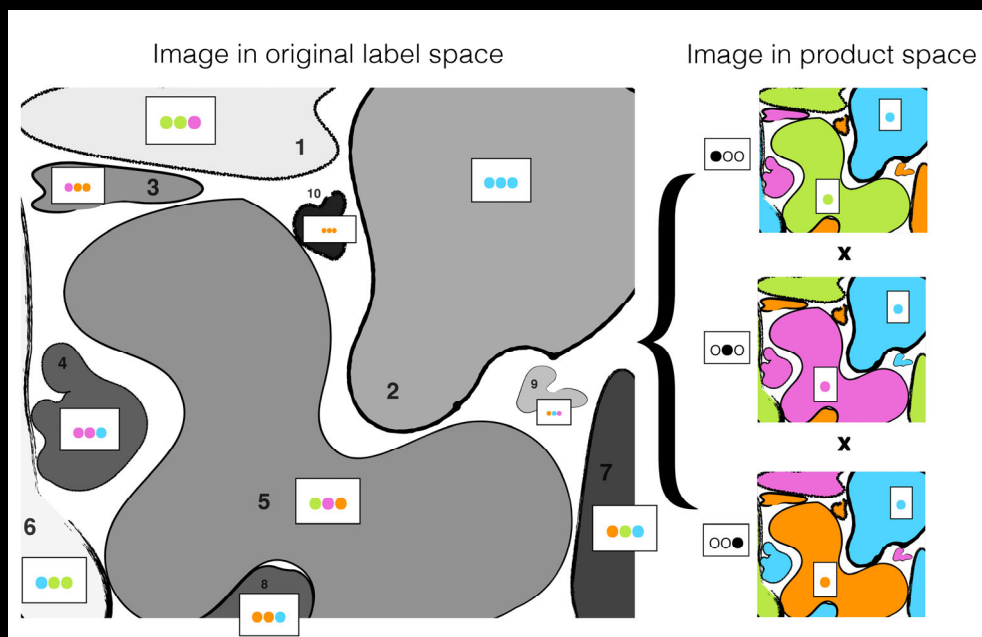


**MaskExtend:** successfully classifies pixels as belonging or not to single neuron

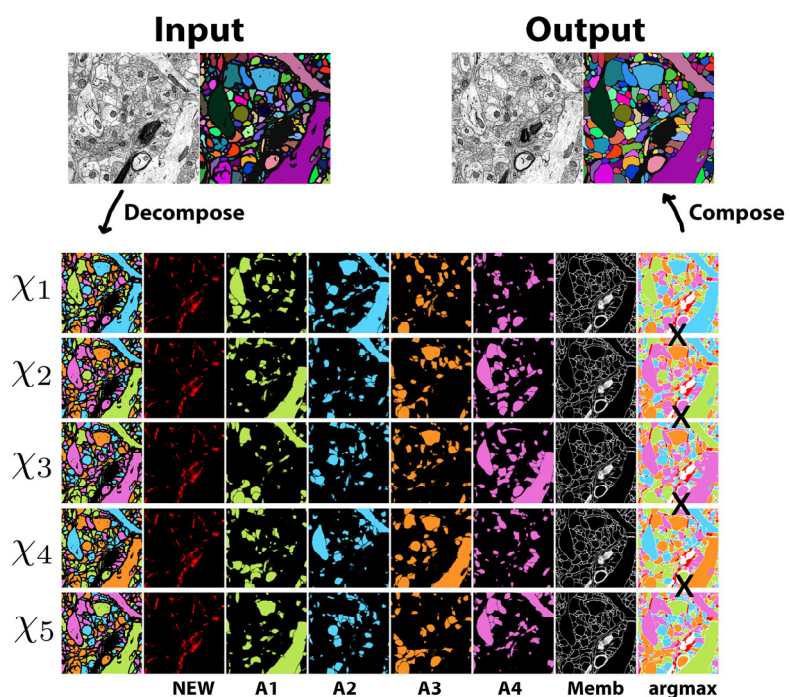


**MultiMask:** cluster pixels into a-priori unknown number of neurons. Even neighbor can change from section to section.

# The 3C Algorithm



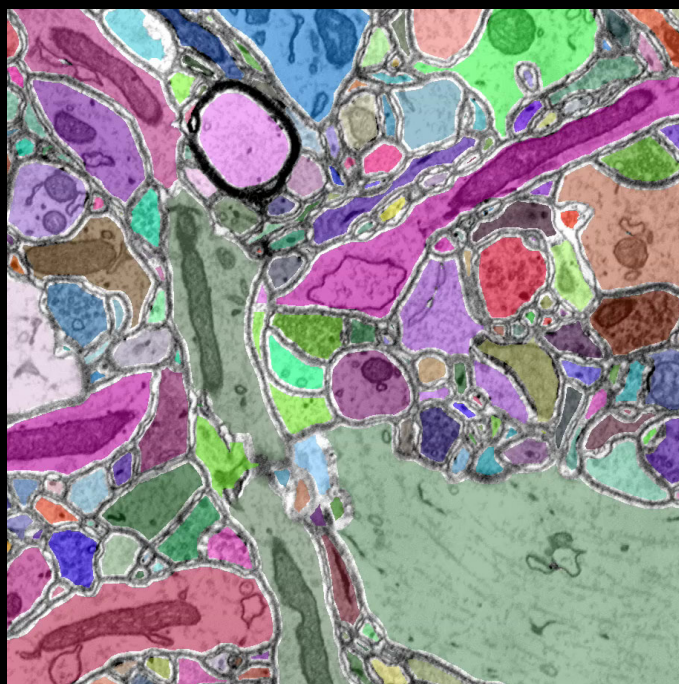
# The 3C Algorithm



Multi  
Mask



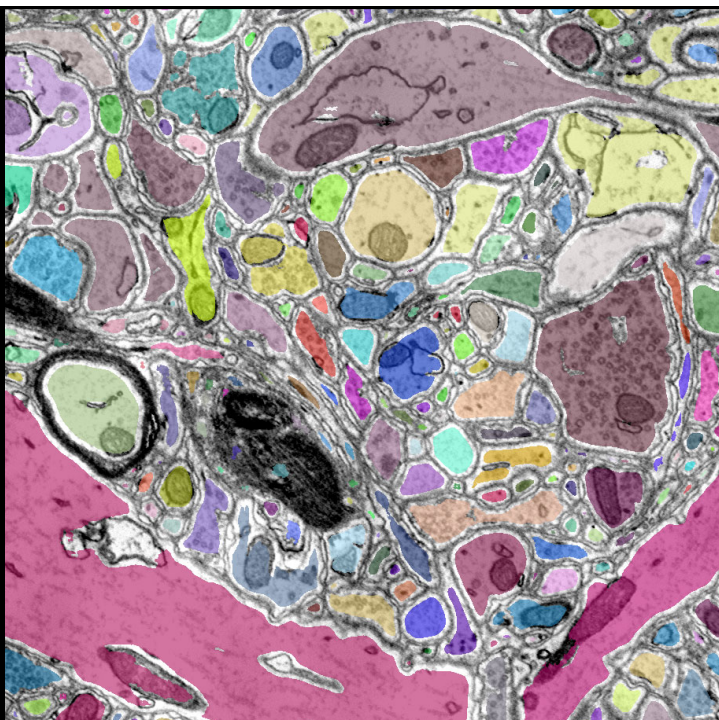
Expansion  
of  
MultiMask  
skeletons  
from  
mouse  
Kasthuri et  
al dataset



Good News:  
Very Few  
Errors

Bad News:  
The  
Errors are  
Harder

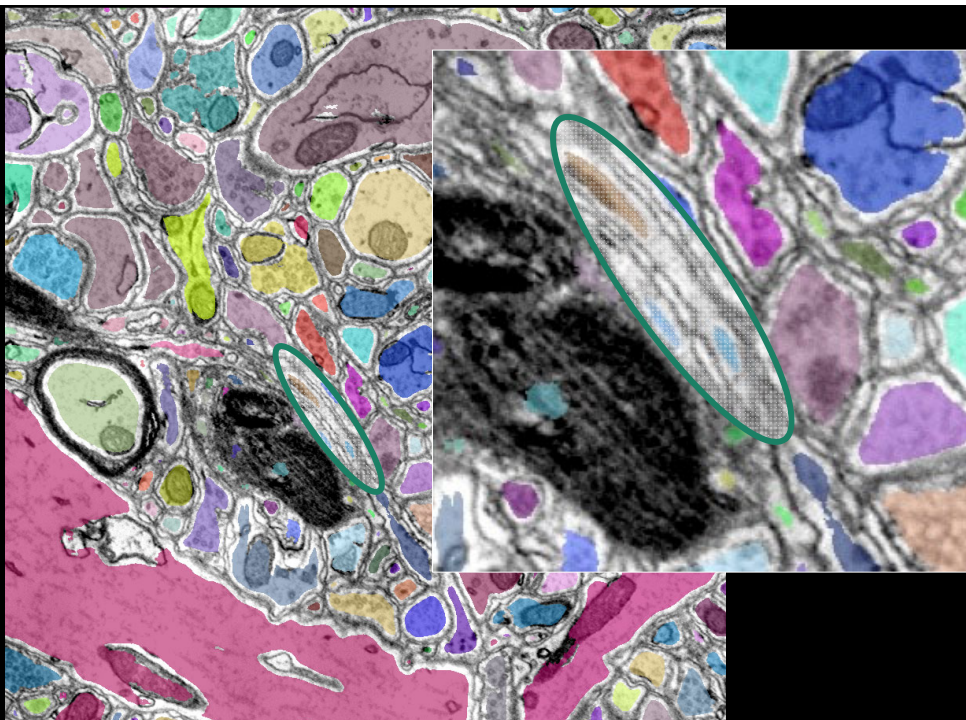
## Spit Errors



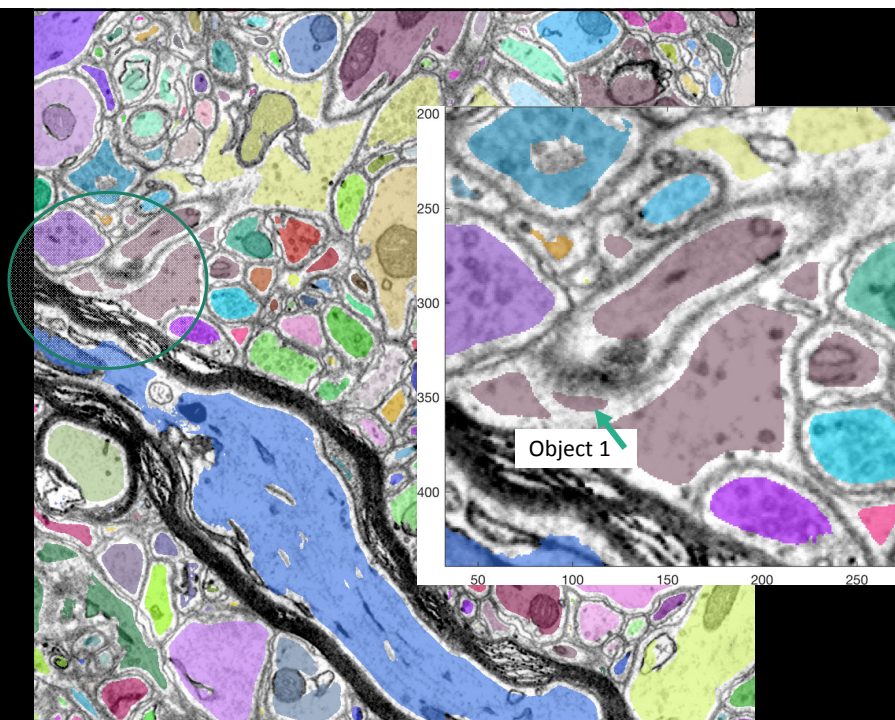
## Spit Errors

Mainly  
objects  
along the  
XY plane...

We are  
working on  
a localized  
solution



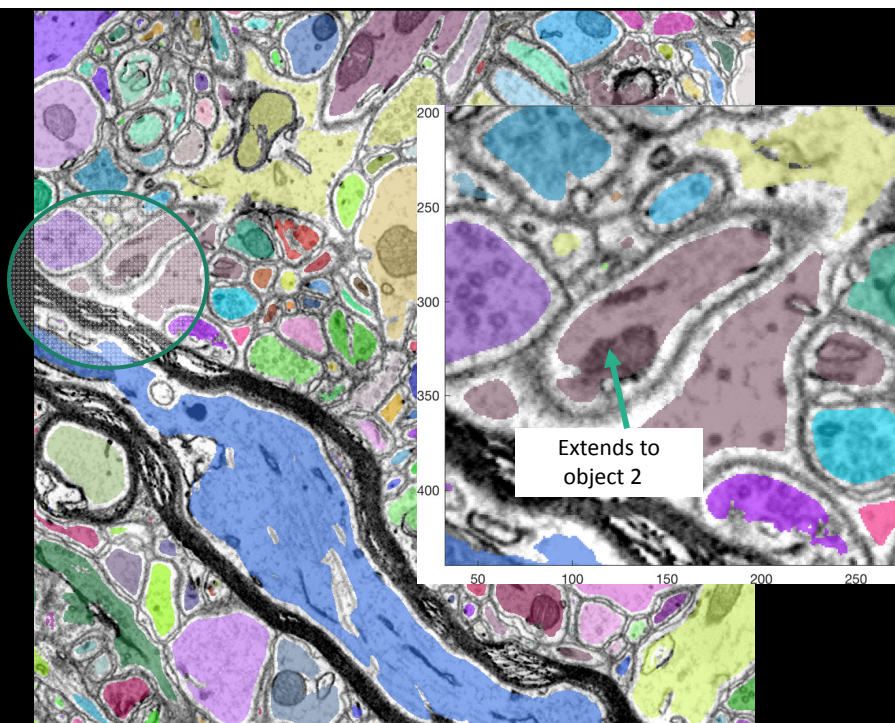
## Merge Errors



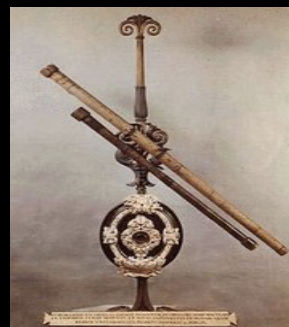
## Merge Errors

Merge errors are due to MultiMask

Merge error detection requires context



# Galileo's Telescope and the future of Connectomics



## The MIT Computational Connectomics Group



Alex Matveev



Yaron Meirovitch



David Budden



David Rolnick



Tim Kaler



Rajeev Parvathala



Hayk Saribekyan



Lu Mi



Shibani Santurkar



Justin Kopinsky