Neuronal Circuit and Synapse Analysis with Deep Neural Networks

Final Presentation

Faezeh Ghazi

Jason Kha

Team

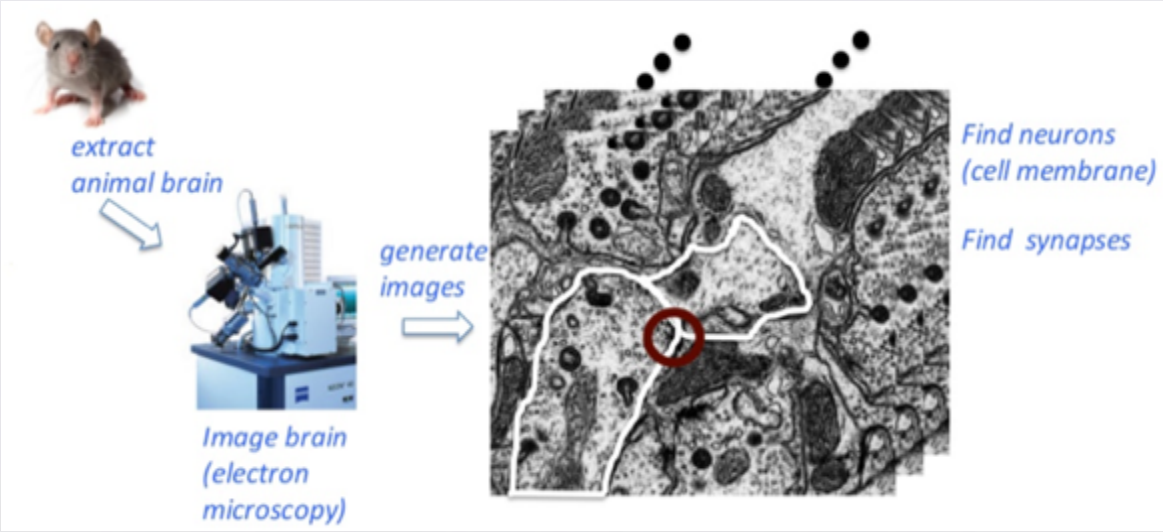
- Dr. Daniela Boassa
  - Director of Core Efforts in Probe Development and Refinement at National Center for Microscopy and Imaging Research (NCMIR)
- Dr. Matthias Haberl
  - Postdoc at NCMIR, Capstone Project Technical Lead

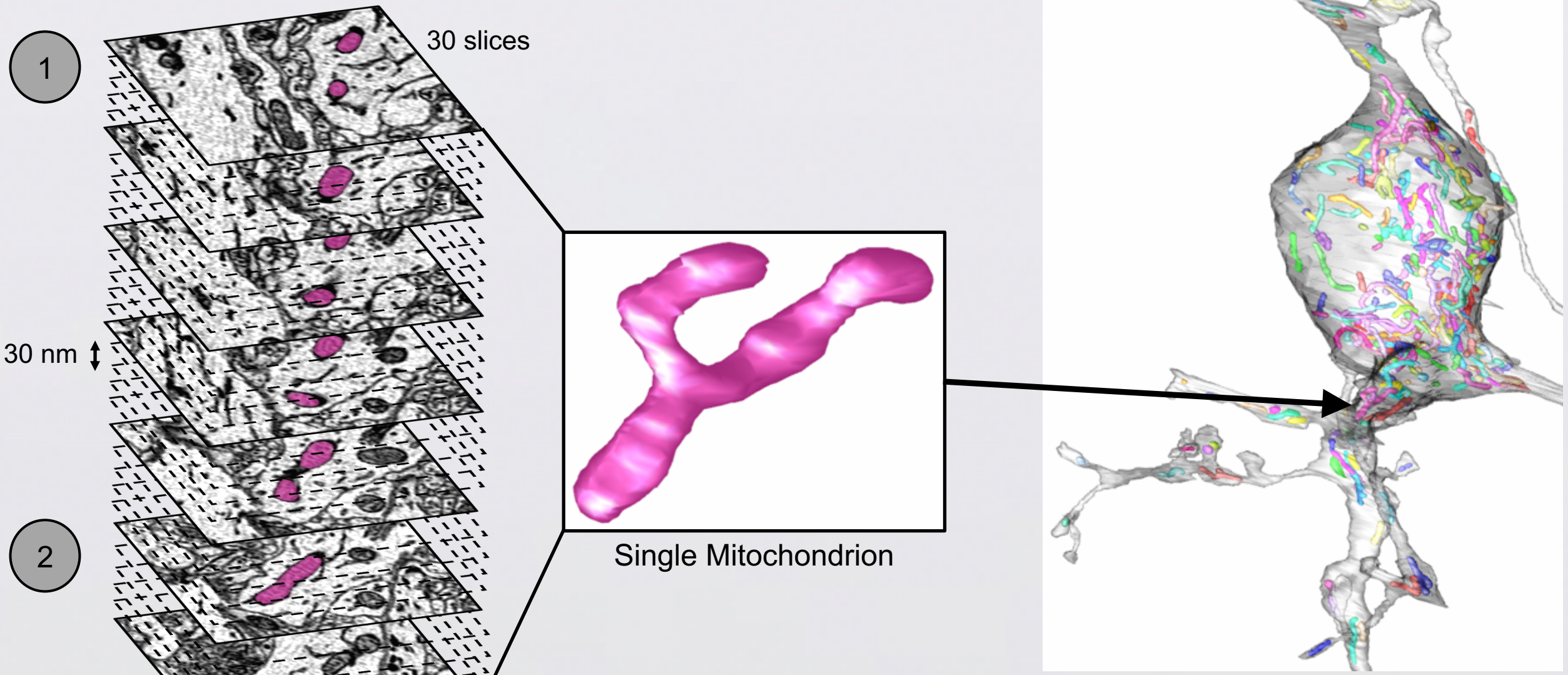
Introduction

- The average human brain has about 86 billion neurons
- Major breakthroughs in neuroscience rely on understanding the structure, function, and connectivity of the brain
- Mapping the brain could help researchers study the onset and progression of neurodegenerative diseases such as Alzheimer's and Parkinson's

How do we map the brain?

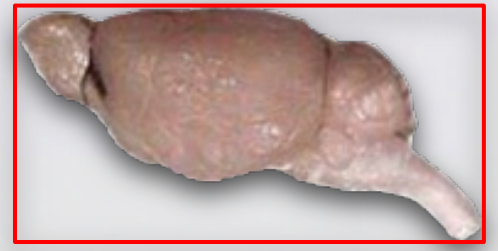
- 3D Image Segmentation
  - Collect brain tissue samples, usually from rodents
  - Perform electron microscopy on brain tissue
  - Locate cell structures in electron microscopy images
  - Create 3D reconstructions of objects




1
30 slices
30 nm
2
Single Mitochondrion

Challenge

- Currently, the most accurate brain image segmentations are annotated manually by humans
- Manual segmentation is a labor-intensive process
- The manual segmentation of all mitochondria in one mouse brain would require 24/7 work for one year from ~2.7 million scientists



~500 mm^3

Problem and Value Statement

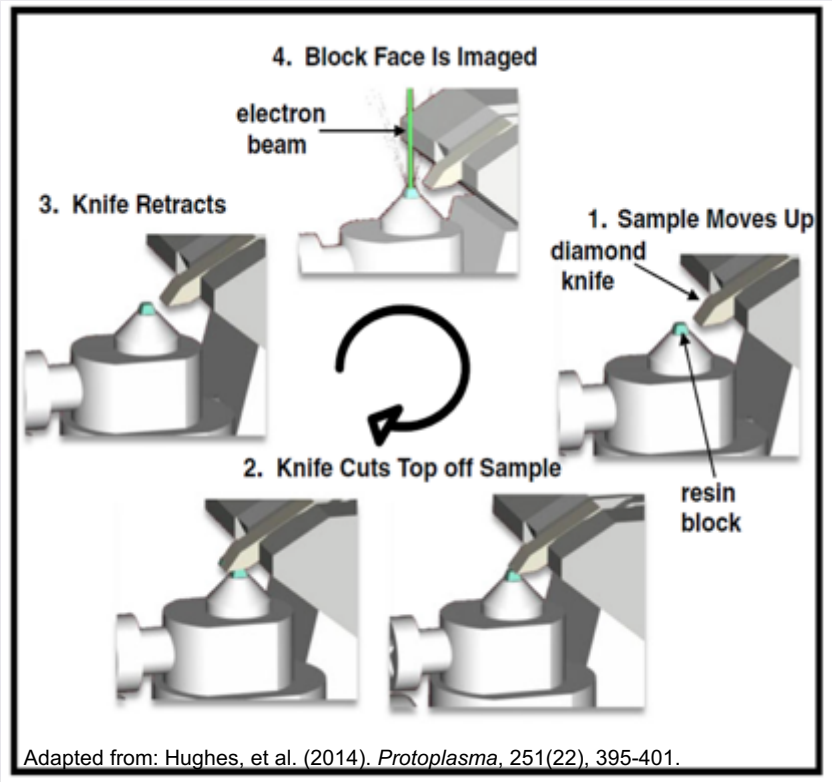
- Manual segmentation of 3D electron microscopy images is unsustainable, but neural networks are capable of automating 3D image segmentation, and therefore can be used to improve the efficiency and accuracy of labeling organelles in brain cells

Our Solution

- CDense3M will improve on the accuracy of automated synapse segmentations compared to the existing CDeep3M tool by using 3D image analyses and flood-filling networks
- Accurate and automated synapse segmentations → faster brain mapping

Data Definitions

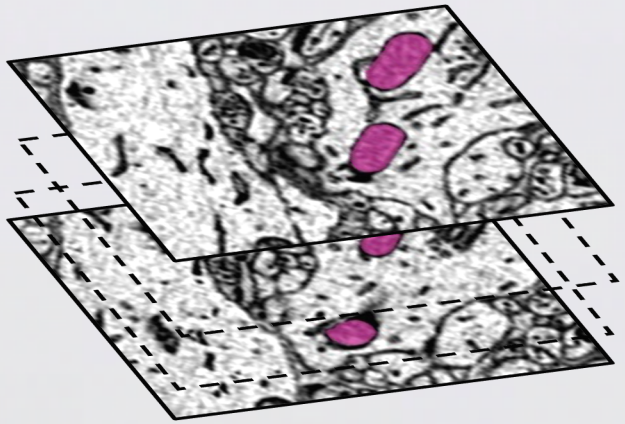
- Serial block-face scanning electron microscopy (SBEM) image volumes with genetically labeled organelles from brain tissue of lab mice
- SBEM constructs high resolution 3D images of tissues by repeatedly using an ultra-thin knife to cut a thin section from a block face and then imaging the next layer



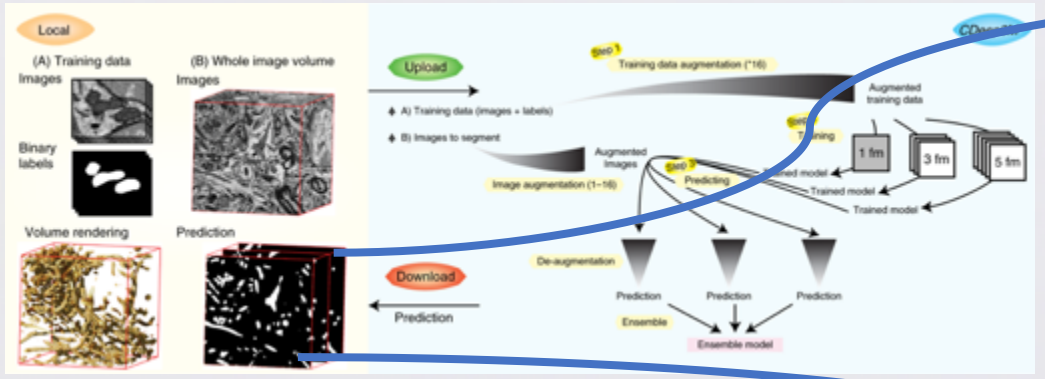
Adapted from: Hughes, et al. (2014). *Protoplasma*, 251(22), 395-401.

Data Format

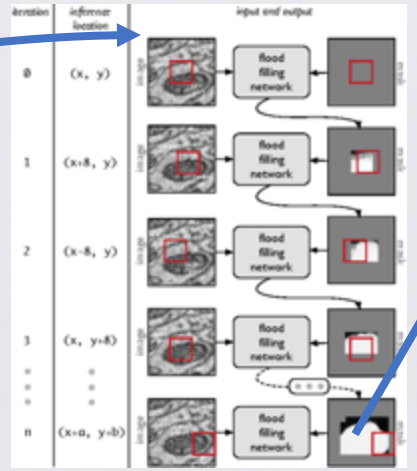
- 1024 x 1024 x 100 voxel image volume of *Nucleus accumbens*
- Stacking the images depth-wise results in a 3D stack of images



Data Pipeline



FFN

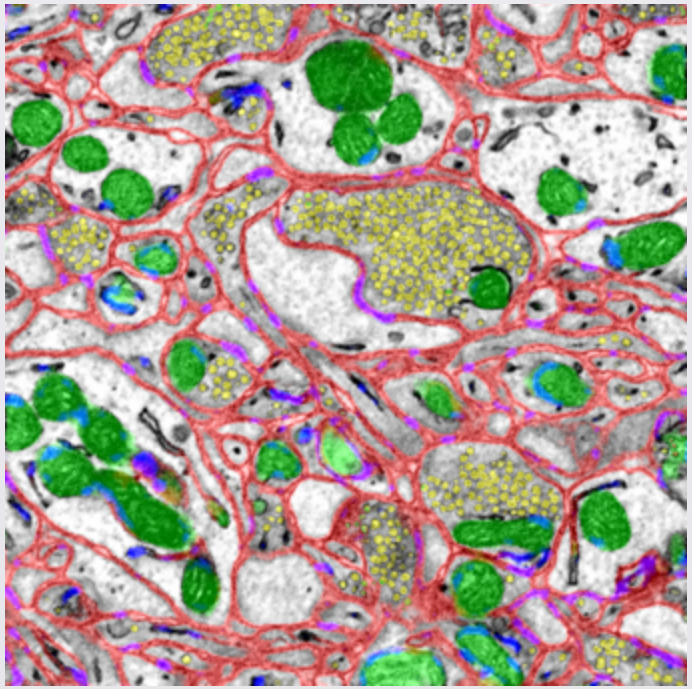


3D Vesicle Filtering

CDense3M

Exploratory Data Analysis

- Membranes and mitochondria segmentations are excellent
- Synaptic vesicles segmentations are good
- Synapse segmentations are poor
  - Many false positives



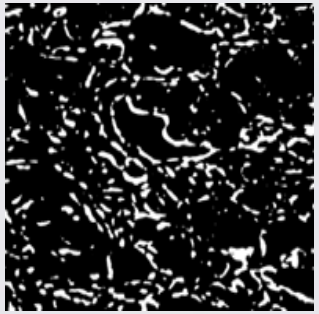
Minimum Viable Modeling Product (MVMP)

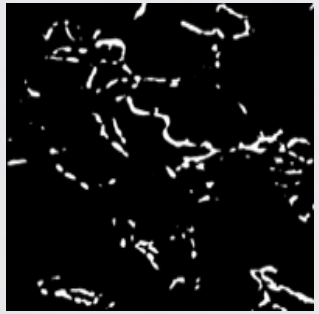
- Based on biological knowledge:
    1. Synapses should **NOT** overlap with mitochondria
    2. Synapses should overlap to a large fraction with the membranes
        1. We set the threshold 80%
- MVMP: Use more accurate segmentations of other organelles to improve synapse segmentations
    - Erase synapses that overlap with mitochondria
    - Erase synapses that do not overlap with membranes
- MVMP results in a 2% increase in the voxel-wise accuracy of synapse segmentations
    - Not good enough – still too many false positives

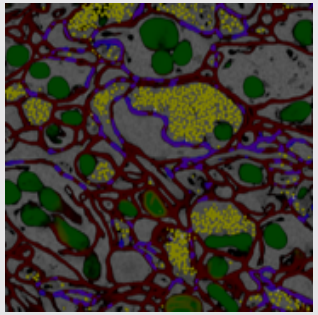
2D Vesicle Filtering

- 3D objects (areas) inside the boundaries may or may not have vesicles
- Set threshold of 5 vesicles and 100 pixels for defining presynaptic area
- Erase those synapses that are not in presynaptic areas

Visualization Results of 2D Filtering







How can we reduce the errors?

Applying 3D Image Processing

- Resolve the problem of 2D filtering
- Resolve the connection of 3D objects of images

3D Image Analyses

- 3D Image Analyses
  - 3D concatenation
  - 3D labeling
  - 3D filtering
    - Erosion Models
    - Determine the threshold
- 3D models improve presynaptic recognition → improved synaptic density predictions

Hypotheses for 3D Filtering and Labeling

- Connection of 3D objects based on the overlapping areas
  - 80% overlapping areas
- In 3D labeling, labeling of new 3D object is related to the labeling of previous 3D objects
- Vesicle erosion can reduce the errors of presynaptic areas' recognition caused by 3D filtering
- Threshold tuning is required for the number and area of vesicles

Create 3D Images, 3D Concatenation, and 3D Labeling

- One dimension is added to all the segmented 2D images, with the images concatenated along the depth of the z-axis
- Concatenated the segmented images of microscopy in the third dimension
- 3D labeling model is calculated both through the similar areas and overlapping areas for the comparison. In both methods, the threshold is 80%.

3D Labeling Algorithms

- Performed labeling for the areas of new images that can be found in the old images
- Performed labeling for the areas of new images that cannot be found in the old images
  - Biggest labeling method: The purpose is to assign the biggest label for those labels of the new images that are not in the labels of previous images

Evaluate 3D Labeling Visualization to Define the Best Model

- Performed visualization based on how much the area of the new image (image 2 for example) is explained by last image (image 1)

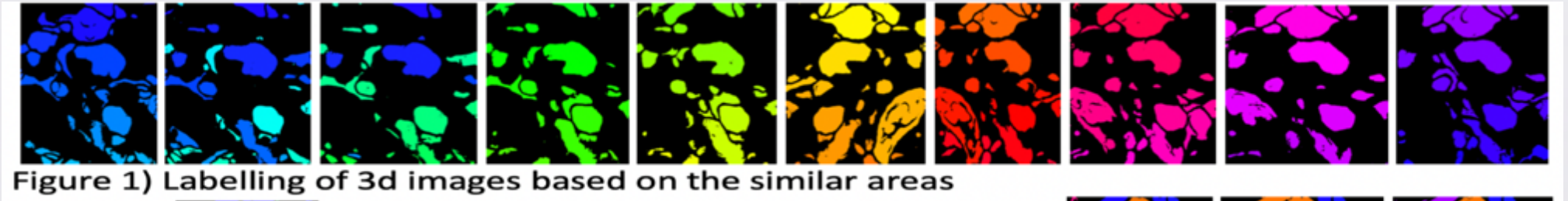


Figure 1) Labelling of 3d images based on the similar areas

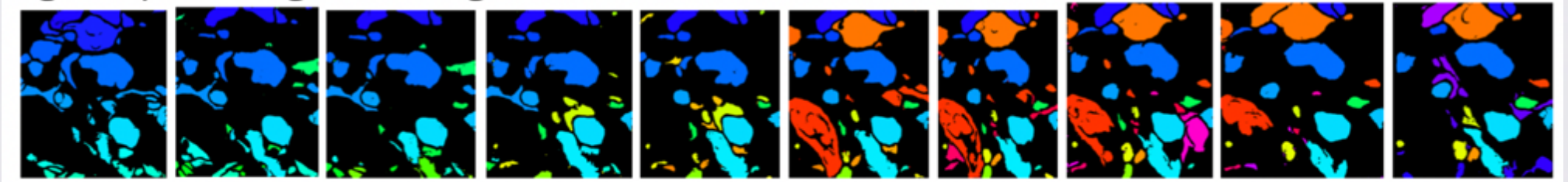


Figure 2) Labelling of 3d images based on the overlapping areas while there are grown vesicles

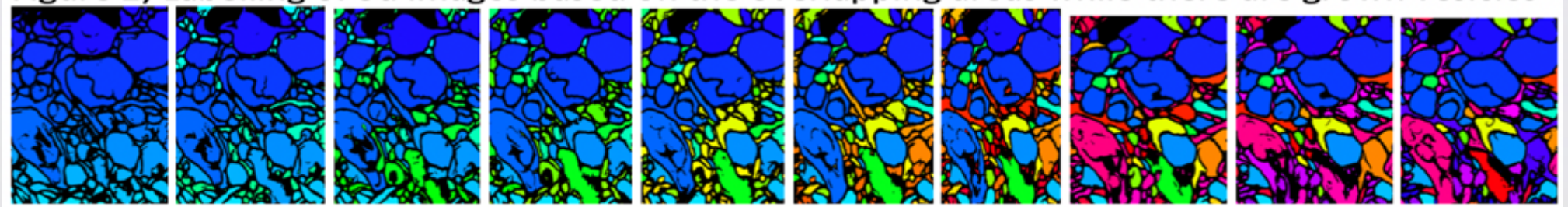


Figure 3) Labelling of 3d images based on overlapping areas

Best 3D Model and its 3D Color Map Labeling Visualization

- The algorithm of this model is to create the random labels from the available labels of our 3D image and then shuffle these random labels

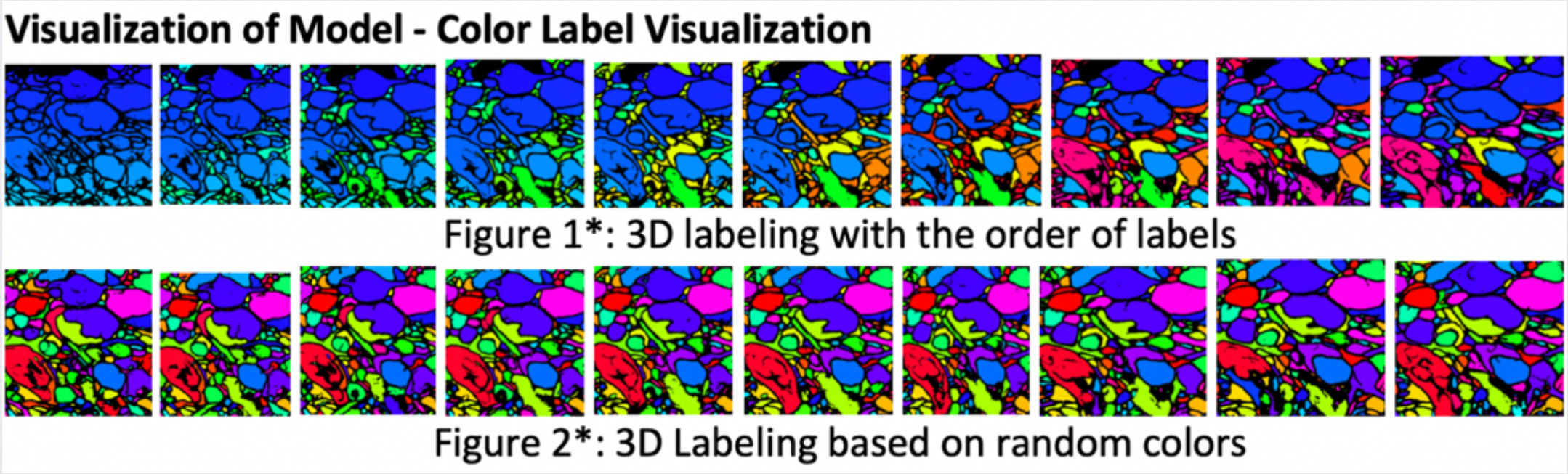


Figure 1*: 3D labeling with the order of labels

Figure 2*: 3D Labeling based on random colors

Modeling of 3D Presynaptic Areas

First Modeling Part

- Apply 3D vesicle filtering with the threshold of 5 vesicles on the 3D label objects. Second, third, and fourth models are related to erosion.

Second Modeling Part

- Focuses on the labeling of the presynaptic site
- Obtained data of this part from the remaining presynaptic sites left from first modeling part
- Obtained labels of presynaptic areas from matching those remaining presynaptic sites with the unique identifiers in our 3D color map

Algorithm Base Comparisons and Accuracy

- Counted the number of vesicles for each label of each image individually
- Received higher accuracy using the individual counting approach. This leads to the improvement of recognizing 3D presynaptic areas based on the 3D vesicles functions.

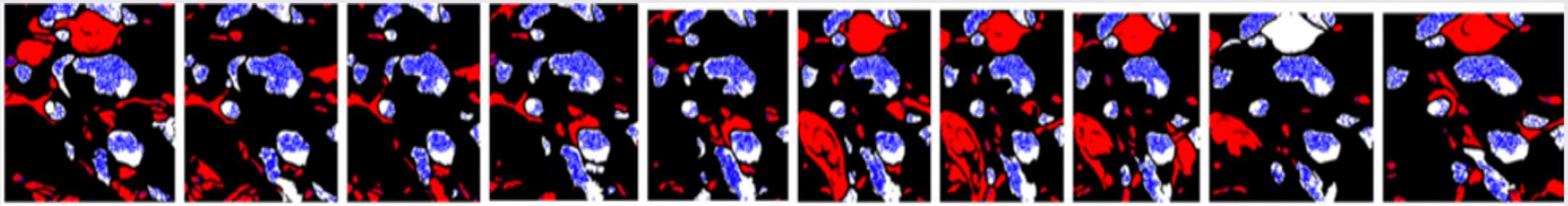
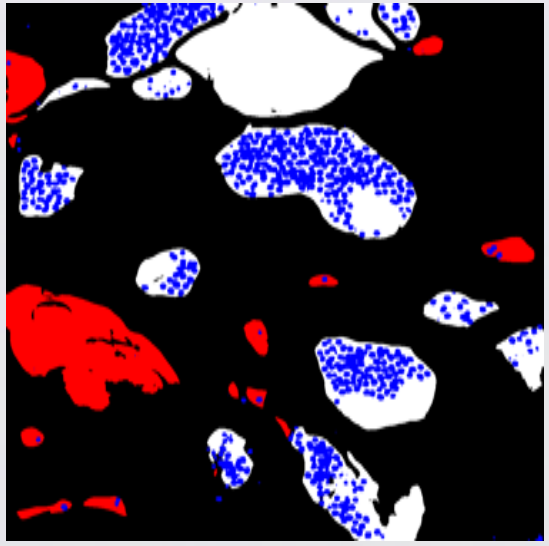


Figure 1) 3D vesicle filtering without erosion

Source of Presynaptic Vesicle Errors

- Dissimilar size of 3D objects throughout the images
- Non-connections of labels in images due to the disconnected boundaries
- Overlapping of vesicles in some labels



Why Erosion Algorithms and Accuracy?

- High error rate for Figure 1 and its label Figure 5 indicates erosion is required
- Applied the erosion algorithms to improve the accuracy of model
  - Defined the kernel to erode vesicles and counted the vesicles before and after the erosion to improve the accuracy of erosion

Visualization of First Modeling Part with Erosion Models

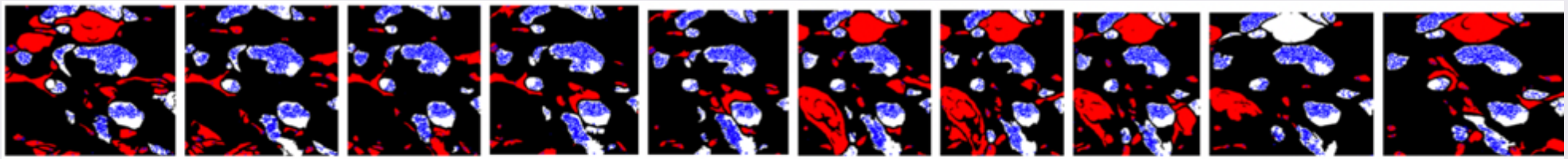


Figure 1) 3D vesicle filtering without erosion

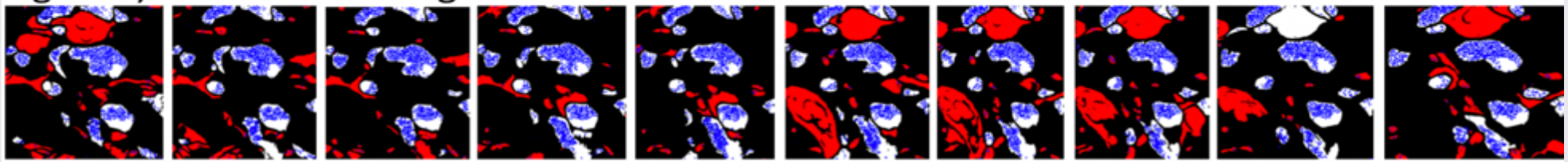


Figure 2) 3D vesicle filtering, applying the erosion of vesicles

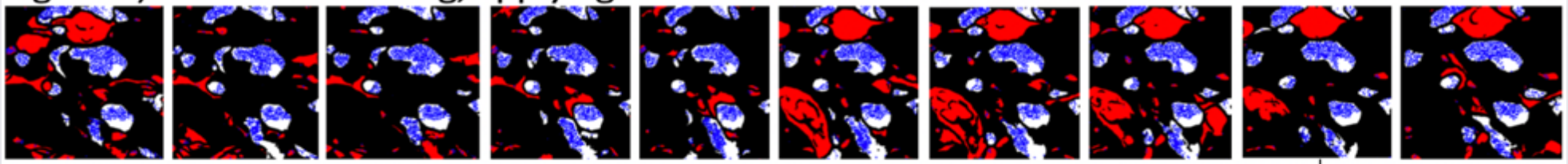


Figure 3) 3D vesicle filtering and applying the erosion of 3D objects (presynaptic)

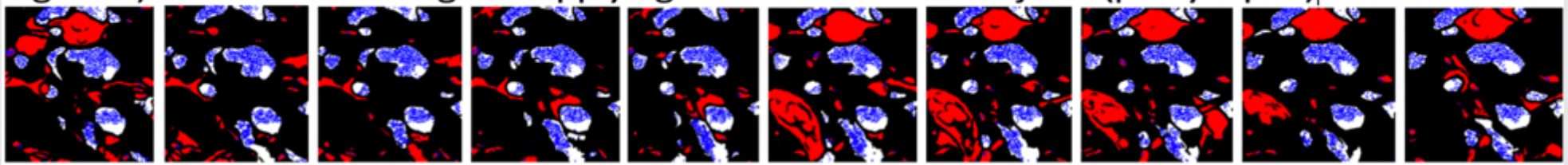


Figure 4) 3D vesicle filtering with two thresholds of 5 vesicles and 500 pixels and applying the erosion of 3D objects (presynaptic)

Visualization of Second Modeling Part

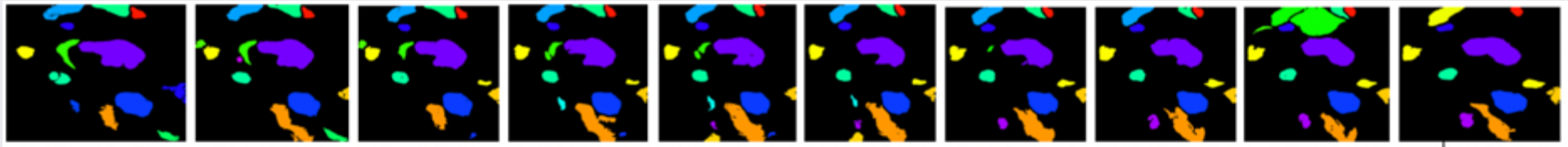


Figure 5) Determination of the labeling of 3D vesicle filtering (Figure 1 presynaptic label)

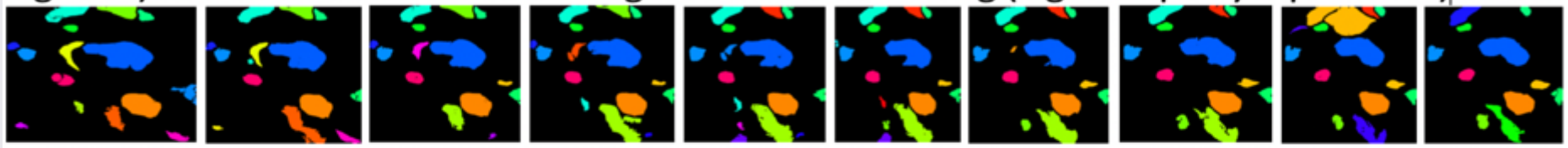


Figure 6) Determination of the labeling of 3D vesicle filtering, applying the vesicle erosion (Figure 2 presynaptic label)

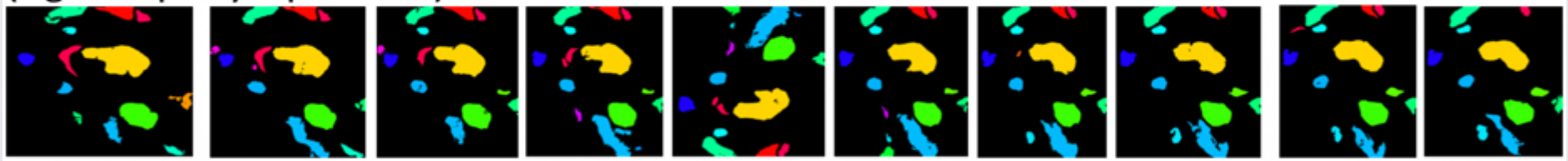


Figure 7) Determination of the labeling of 3D vesicle filtering, applying the 3D objects (presynaptic) erosion (Figures 3 presynaptic label)

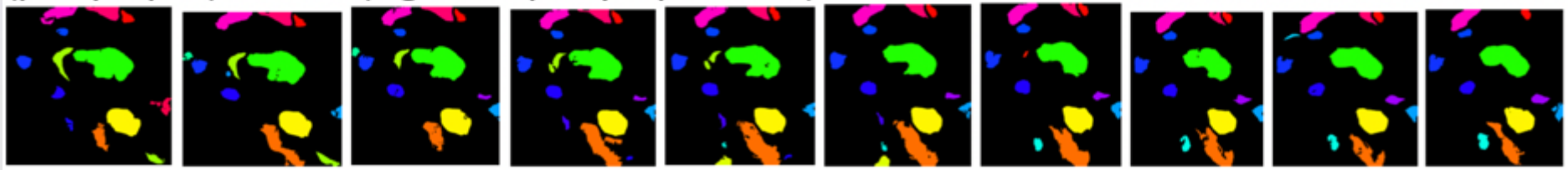


Figure 8) Determination of the labeling of 3D vesicle filtering with two threshold that are 5 vesicles and 500 pixels along with 3D objects (presynaptic) erosion. (Figure 4 presynaptic label)

Interpretation of Visualized Models

Vesicle Filtering Solution Models

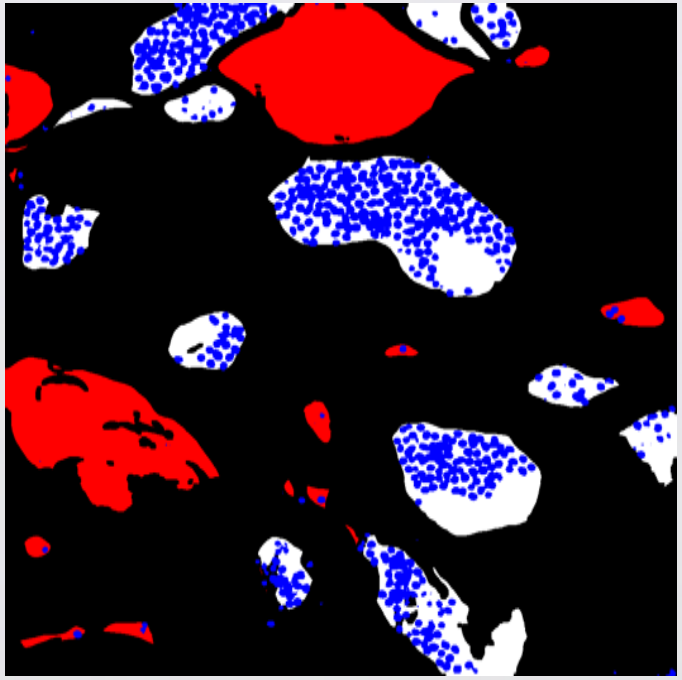
- Figure 2 and its label Figure 6, which is the implementation of vesicle erosion, caused an increase in errors

- Figure 3 and its label Figure 7, which is the implementation of 3D objects (presynaptic) erosion, caused the reduction of errors

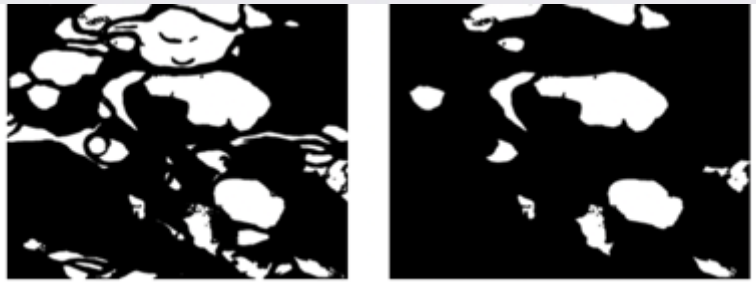
- Figure 4 and its label Figure 8, which is the implementation of applying 3D objects (presynaptic) erosion, along with setting two thresholds for 3D vesicle filtering that are 5 vesicles and 500 pixels, reduces the error the most

Effects of Best 3D Filtering on the Errors

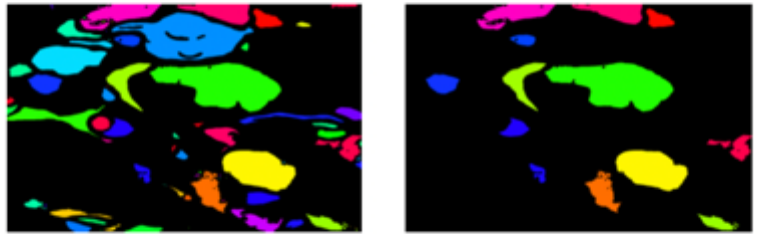
- Combination of 3D filtering with two thresholds of 3D filtering, which are 5 vesicles with 500 pixels, along with presynaptic erosion, reduces the errors



Comparison of Presynaptic Areas Before and After the Best Filtering Model



Binary images of presynaptic areas before filtering on the left side and after vesicle filtering of labeling with threshold 5 and 500 along with 3D objects erosion (best selected modeling part 1) on the right side

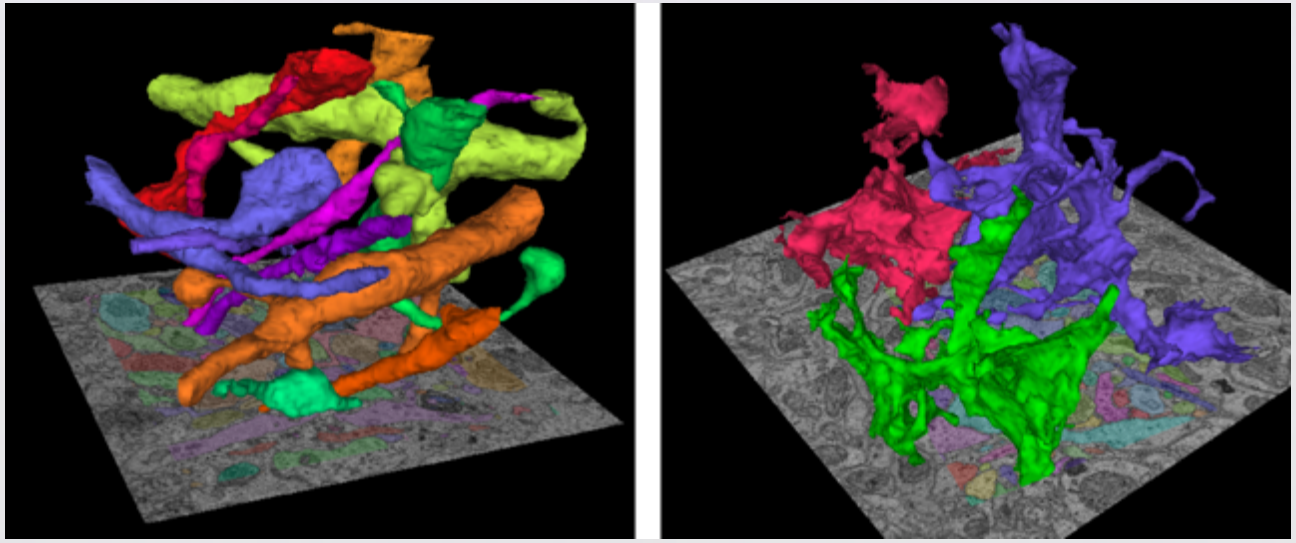


Labeling of binary images of presynaptic areas before filtering on the left side and after vesicle filtering of labeling with threshold 5 and 500 along with 3D objects erosion (best selected modeling part 1) on the right side

3D Image Analyses Demo

Flood-Filling Networks (FFN)

- Trace brain cells in SBEM data
- Recurrent 3D convolutional network that directly produces individual segments from image
- Use FFN 3D cell-body segmentations to improve CDense3M's synapse segmentations



FFN Models

FIB-25

- Trained on FIB-25 - medulla from fruit fly optic lobe
- 4 minutes to segment 250x250x100 image volume
  - 1.6 million voxels/minute

Membrane Prediction

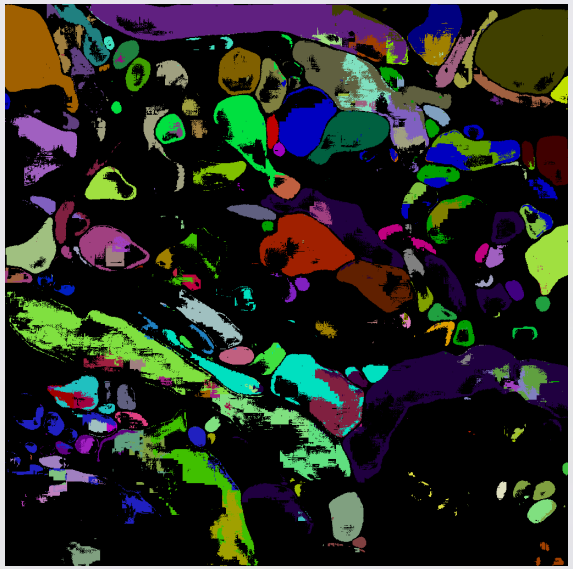
- Trained on CDeep3M membrane segmentations
- 64 minutes to segment 1024x1024x100 image volume
  - 1.6 million voxels/minute

FFN Hypothesis

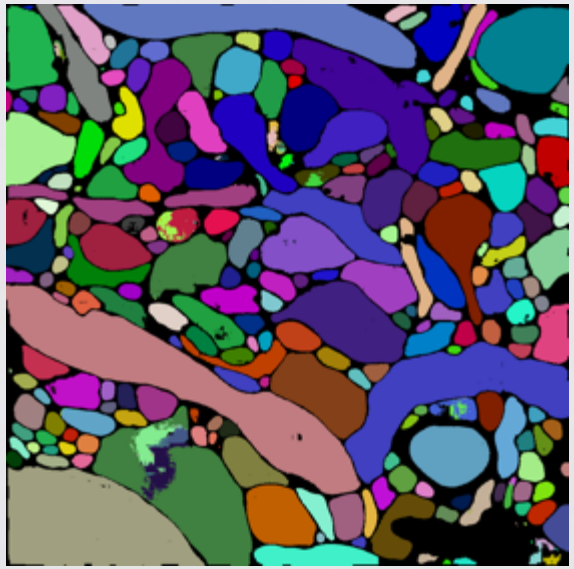
- An FFN model trained on membrane segmentations is more generalizable than the FIB-25 FFN model trained on the optic lobe of *Drosophila* (fruit fly) on different types of electron microscopy (EM) data
  - CDense3M membrane segmentations are 8-bit grayscale images, which are similar in structure regardless of the raw EM data source
  - The raw EM data from the brain of a fruit fly may look slightly different than that of a mouse, which could make transfer learning less effective

FFN Findings

- The provided FFN membrane model checkpoint at 6 million epochs did not cleanly segment even the FFN training data
- But retraining the model for another 2 million epochs resulted in an excellent segmentation on the training data



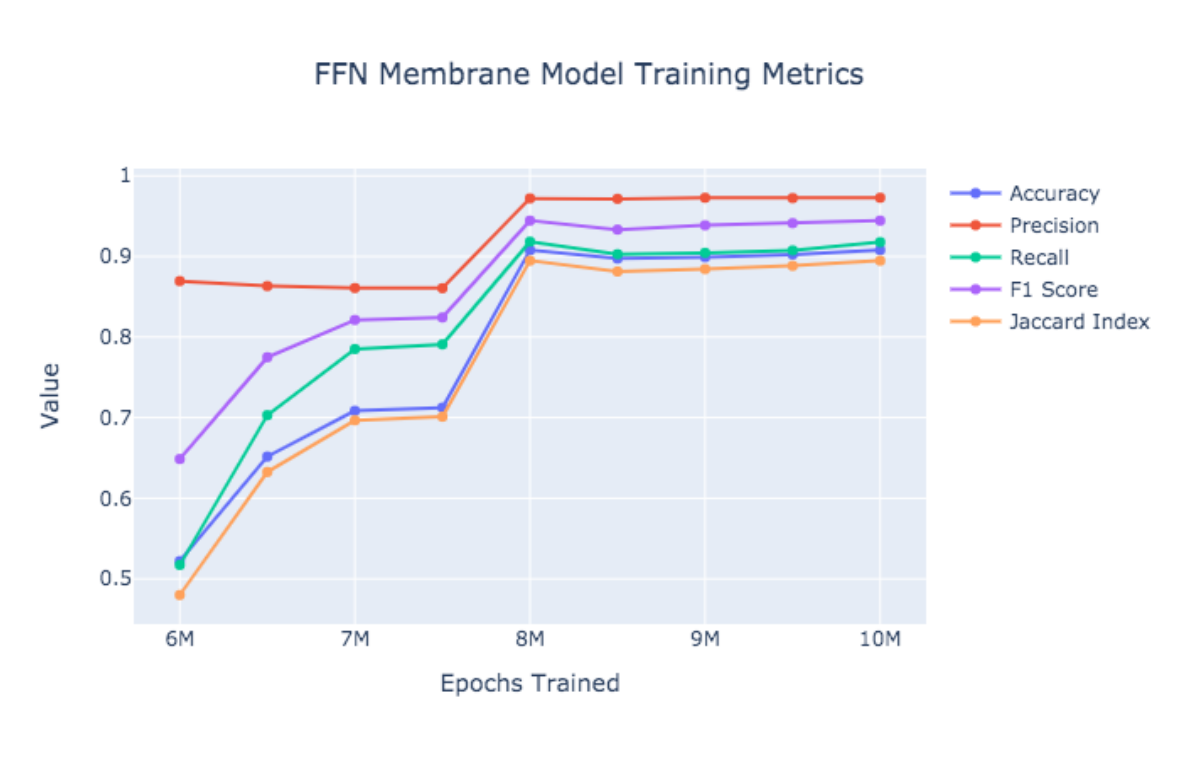
6 million epochs



8 million epochs

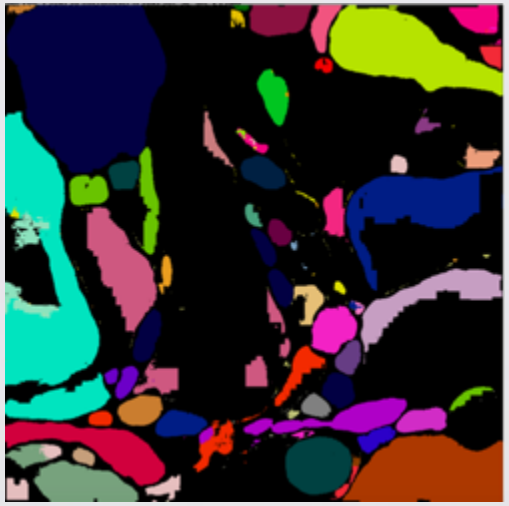
FFN Findings

- Continued training to 10 million epochs, but training metrics did not improve after 8 million epochs
- Use 8 million epochs checkpoint as model to avoid overfitting

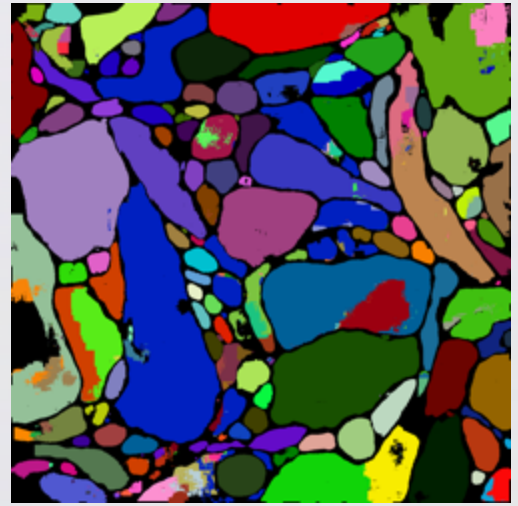


FFN Findings

- The FFN membrane model expects the membrane segmentations input to closely match the *xy* dimensions of the segmentations it was trained on
- Otherwise, segmentation quality will suffer and be distorted



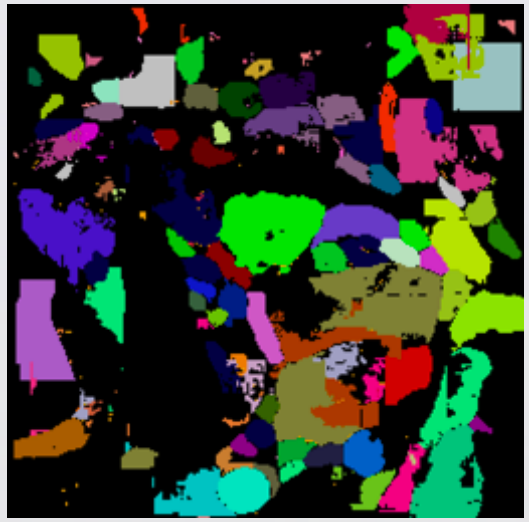
Unscaled



Scaled

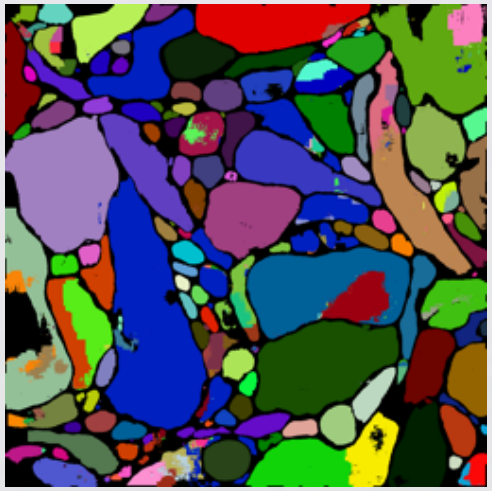
FIB-25 Results

- FIB-25 model performs poorly on *Nucleus accumbens* data
- Fails to locate seed points in all the black areas
- Missing many of the smaller cells
- Decent precision, low recall



Membrane Model Results

- Membrane model is more accurate than FIB-25 on *Nucleus accumbens* data
- Better seeds → Better segmentations
- Some holes due to mitochondria being mistakenly segmented as membranes by CDense3M
- Supports our hypothesis that the FFN membrane model is more generalizable than FIB-25 on new data



Evaluating FFN Segmentation Accuracy

- Measure per voxel (3D pixel)
- One-hot encode each cell and apply binary segmentation mask for each cell ID
  - 1: voxel is part of cell
  - 0: voxel is not part of cell
- Count number of voxels that are true positives, true negatives, false positives, and false negatives for each cell ID using the predicted and ground truth segmentations
- Take the average value among all cells for each metric

FFN Metrics

	Training Data with Membrane Model	Test Data with Membrane Model	Test Data with FIB-25 Model
Accuracy	0.9078	0.8237	0.6651
Precision	0.9720	0.9109	0.8737
Recall	0.9184	0.8262	0.7193
F1-score	0.9444	0.8665	0.7890
Jaccard index	0.8947	0.8041	0.6464

FFN Scalability

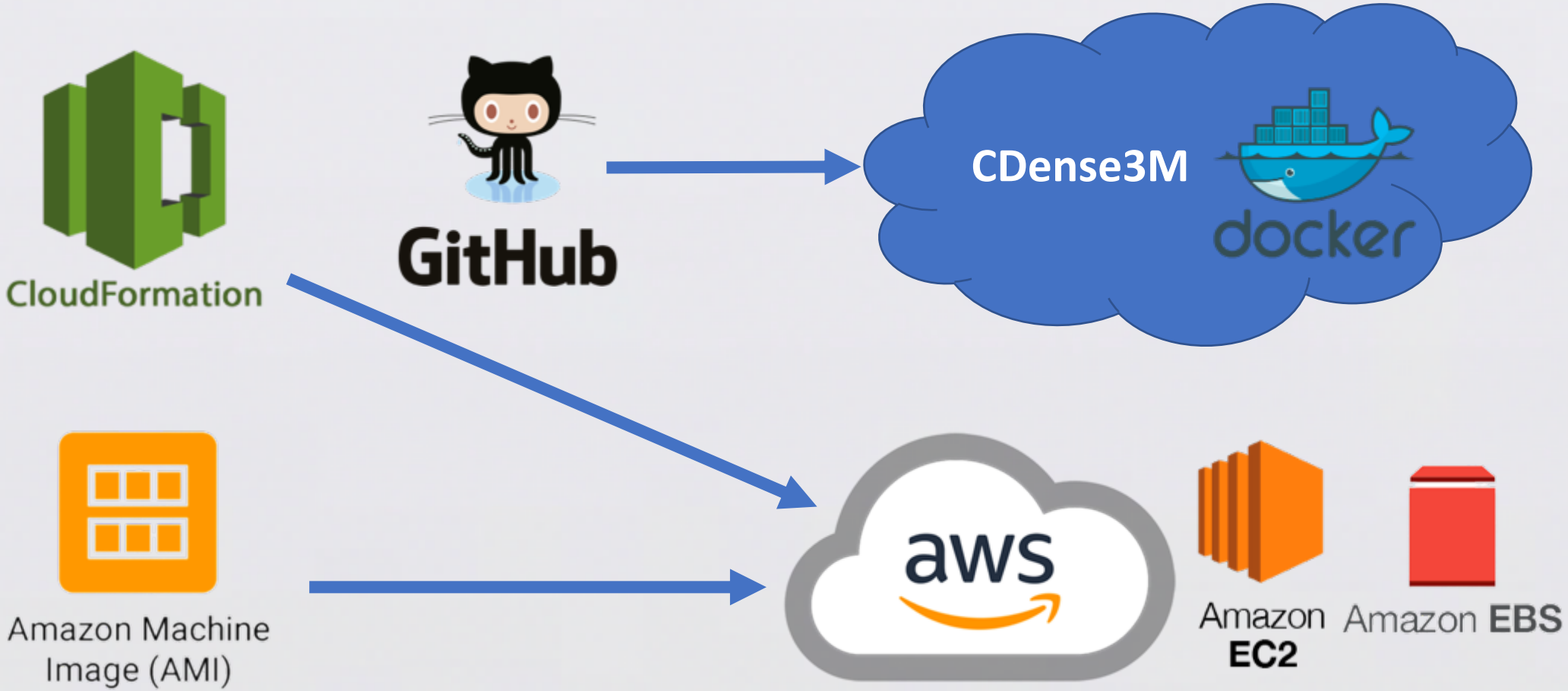
- Both FFN membrane and FIB-25 models run in linear time with respect to the total number of voxels
  - Example for 16 GB of GPU RAM (AWS g4dn.xlarge EC2 instance)
    - 1024x1024x100 voxels takes 64 minutes to segment
    - 750x750x100 voxels takes 31 minutes to segment
    - 1.6 million voxels/minute
  - But when we double the GPU RAM to 32 GB (g4dn.2xlarge)
    - 1024x1024x100 voxels takes 40 minutes to segment
    - 750x750x100 voxels takes 22 minutes to segment
    - 2.6 million voxels/minute
  - CDense3M is 50% more efficient when doubling GPU RAM to 32 GB
    - Certain stages of the pipeline run jobs in parallel

Business Value of Findings

- 10% improvement in the accuracy of automated synapse segmentations using 3D modeling and FFN → 10% fewer mistakes that a human expert needs to manually correct
- FFN membrane model generalizability makes it one-size-fits-all
  - Transfer learning = Save time and money by not having to train own model
- Results are reproducible
  - CDense3M extends CDeep3M
  - No need to install or learn new stack like the similar Janelia DVID tool
  - Command to segment images (runprediction.sh) in CDense3M is very similar to CDeep3M, with the only difference being two new options for running FFN in CDense3M

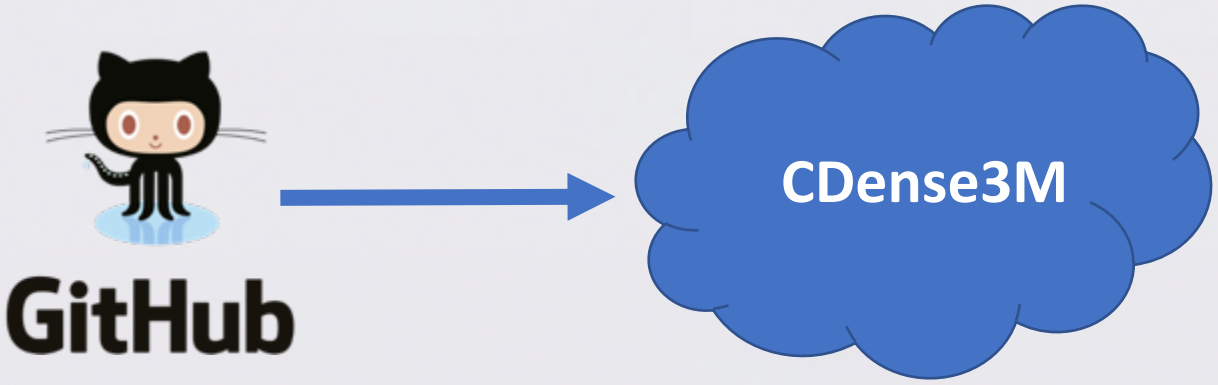
How can I start using CDense3M?

Solution Architecture - Production



Cost: $0.54/hour for one g4dn.xlarge EC2 instance + 100 GB EBS volume

Solution Architecture - Development



colab

Google Drive

Cost: FREE

Disclaimer: Colab resources are not guaranteed and not unlimited, and the usage limits sometimes fluctuate. CDense3M shall not be liable for any loss or damage to property caused by Colab's usage limits.

Business Value of Solution Architecture

- Supports multiple environments
  - AWS
  - Colab
  - On-premises
- Easy installation
  - Less time spent troubleshooting, more time for research
- Accessibility – caters to different types of users
  - CLI for advanced users
  - GUI for beginners
- Free option – albeit with limits

Next Steps

- Calculate more accurate threshold for 3D labeling and vesicle filtering
- Clean up CDense3M membrane segmentations
  - False mitochondria in membrane segmentations (confused with myelin sheaths) reduces the cell-body segmentation accuracy of the FFN membrane model
- Crop large images into smaller components and process them individually to handle memory constraints
- Publish a paper on CDense3M after implementing these improvements

FFN Demo

Conclusion

- All the hypotheses at the beginning were confirmed
- Verified the best 3D labeling is based on the overlapping of 3D objects with a threshold of over 80%
- Verified the best 3D vesicle filtering is based on keeping the 3D labeling overlapping relation with the individual vesicles analyses in each image. This results in a higher accuracy.
- All the data of 3D image processing in the next step is relevant to the previous step
- The threshold of our 3D filtering for our best model is 5 vesicles and 500 pixels

Conclusion

- CDense3M achieves a 10% increase in the segmentation of synapses compared to CDeep3M2
- CDense3M is easy to install and use
- The ultimate goal of CDense3M is to help researchers accelerate their study of neurodegenerative diseases like Alzheimer's and Parkinson's disease in the search of better treatment options, or maybe even a cure

Acknowledgements

- Dr. Daniela Boassa and Dr. Matthias Haberl
- Dr. İlkay Altıntaş
- Dr. Mark Ellisman
- NCMIR staff
- Our families
- Our employers
- DSE MAS faculty
- DSE MAS staff
- Our classmates

Thanks for listening!

Questions?