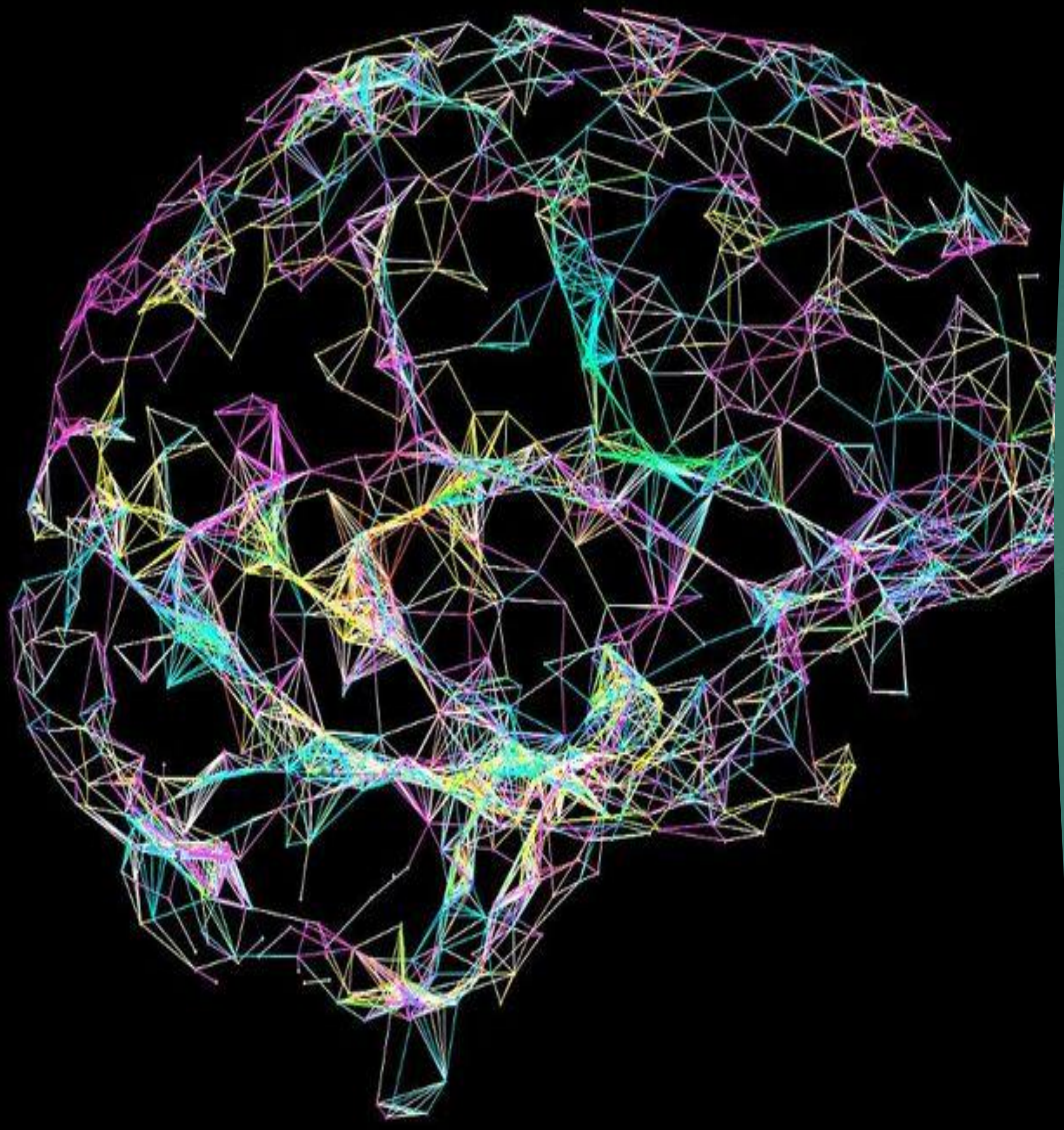


ALZHEIMER'S DISEASE DATA ANALYSIS

Group 3, June 05, 2020

Team Members:

Ashok Mondal

Daniel Silva

Doreh Kazemisefat

Matthew Bond

Peishan Sha

Advisors:

Dr. Mark Ellisman

Matthew Madany

Steve Peltier

Project Overview

- Existing System
  - Environment deployed on Nautilus
  - CDeep3M for image segmentation
- Goal
  - Improve original image quality
  - Enhance model performance
  - Improve output visualizations of brain organelles
  - Improve image segmentation and volume rendering

Approach

- Understanding Challenges
- Data Acquisition and Pipeline
  - Execute in cluster environment
  - Pull models and dependencies without user interaction
- Exploratory Data Analysis
- Define Hypothesis
  - Develop/Validate Hypothesis
- Solution Engineering
  - Architecturing
  - Develop model
  - Validate results and optimization
  - Final Product

Challenges

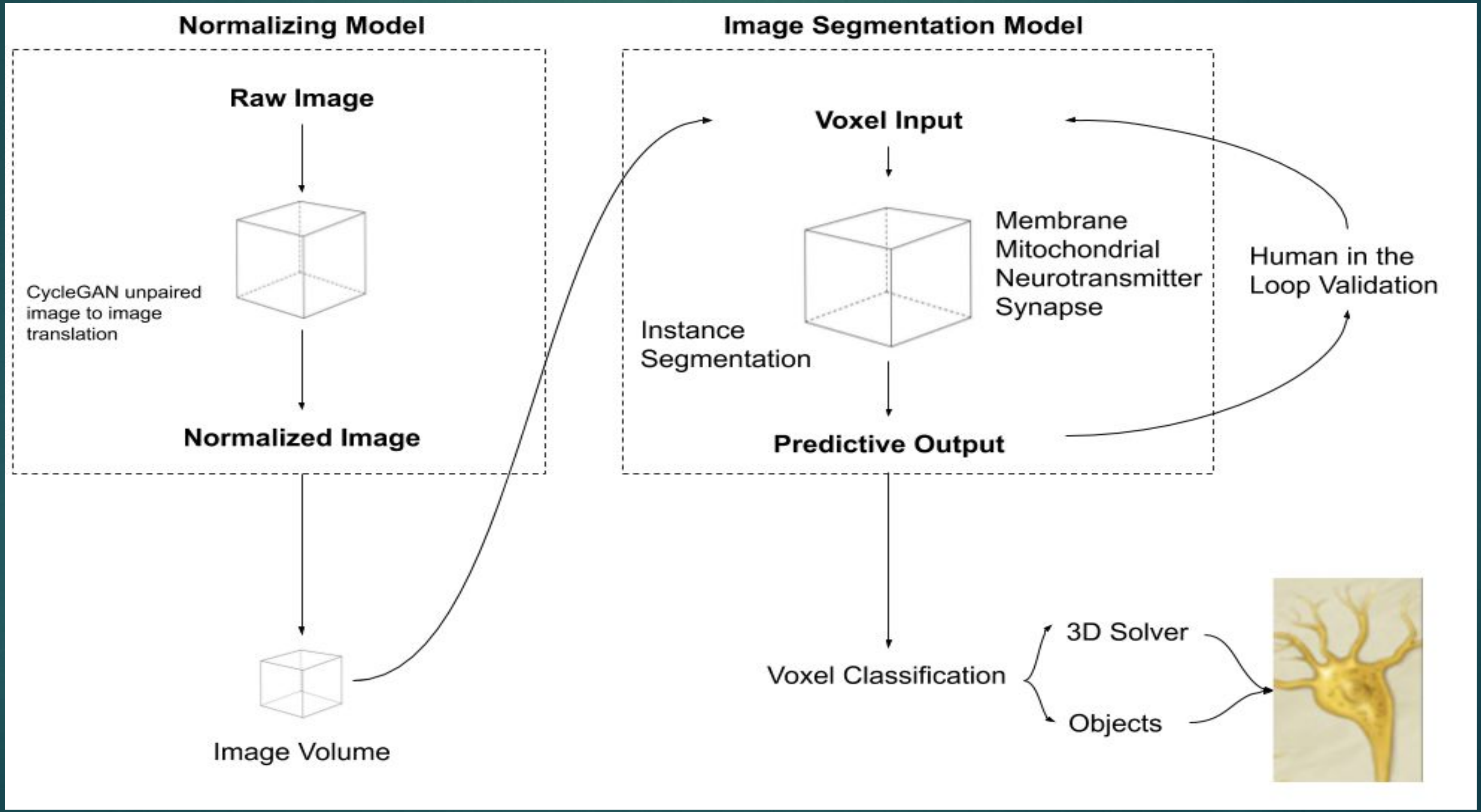
- Setting up the environment and installing required libraries
- Low quality image data from Electron Microscope
- Scale of data makes analysis of any individual data set extremely difficult without dedicated supercomputer resource (e.g. 3500 x 3500 x n pixel image stacks)
- Persons with high levels of understanding were required to label every feature of each cell by hand to generate ground truth data

Data Acquisition

Four datasets from different brain areas of mice provided by NCMIR

Database Name	Source Location	Destination in Data Pipeline	Data Movement and Processing Scripts and Notebooks	Data Size
Cell Image Library (Public) -SBEM -TEM	http://www.cellimagelibrary.org/cdeep3m		Jupyter Notebook	100 GB
Cerebellum	Google Drive	As Target domain data in CycleGAN Process	Jupyter Notebook	2.92 GB
Cortext_1	Google Drive	- Original images as Source domain in CycleGAN Process - Generated Images as CDeep3M input for image segmentation	Jupyter Notebook	3.5 GB
Cortext_2	Google Drive	“ ”	Jupyter Notebook	7.1 GB
Hypothalamus	Google Drive	“ ”	Jupyter Notebook	1.3 GB

Data Pipeline



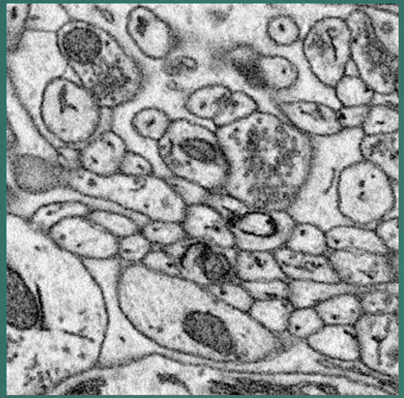
Exploratory Data Analysis

- Understanding Data from different part of brain
  - Cortex and Hypothalamus data is of poor quality, resulting in poor instance segmentation
  - Cerebellum data is of higher quality
- Understanding Cycle-Consistent Generative Adversarial Networks (CycleGAN)
  - Improve low quality image through CycleGAN process

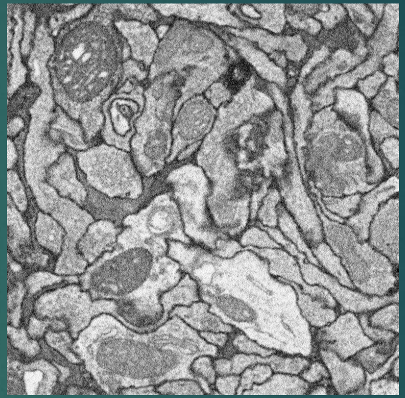
Data Sample

- Electron microscopy samples from mouse brain
- Pixel-level details of nanoscale structures
- The capability of imaging millimeter, micrometer, and smaller sizes
- 2D images, the length of the Z axis is simply the depth of the image stack

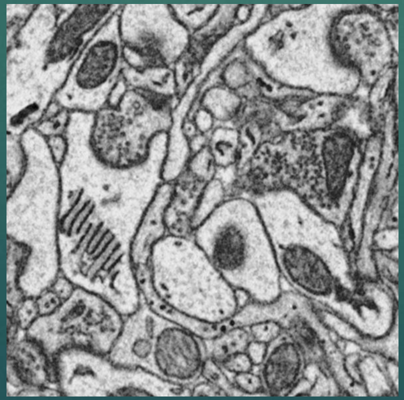
Cerebellum



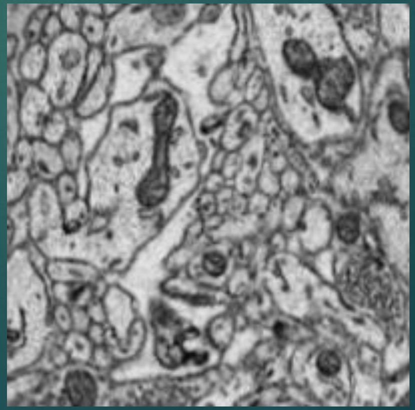
Cortex_2



Cortex_1



Hypothalamus



Data Preparation

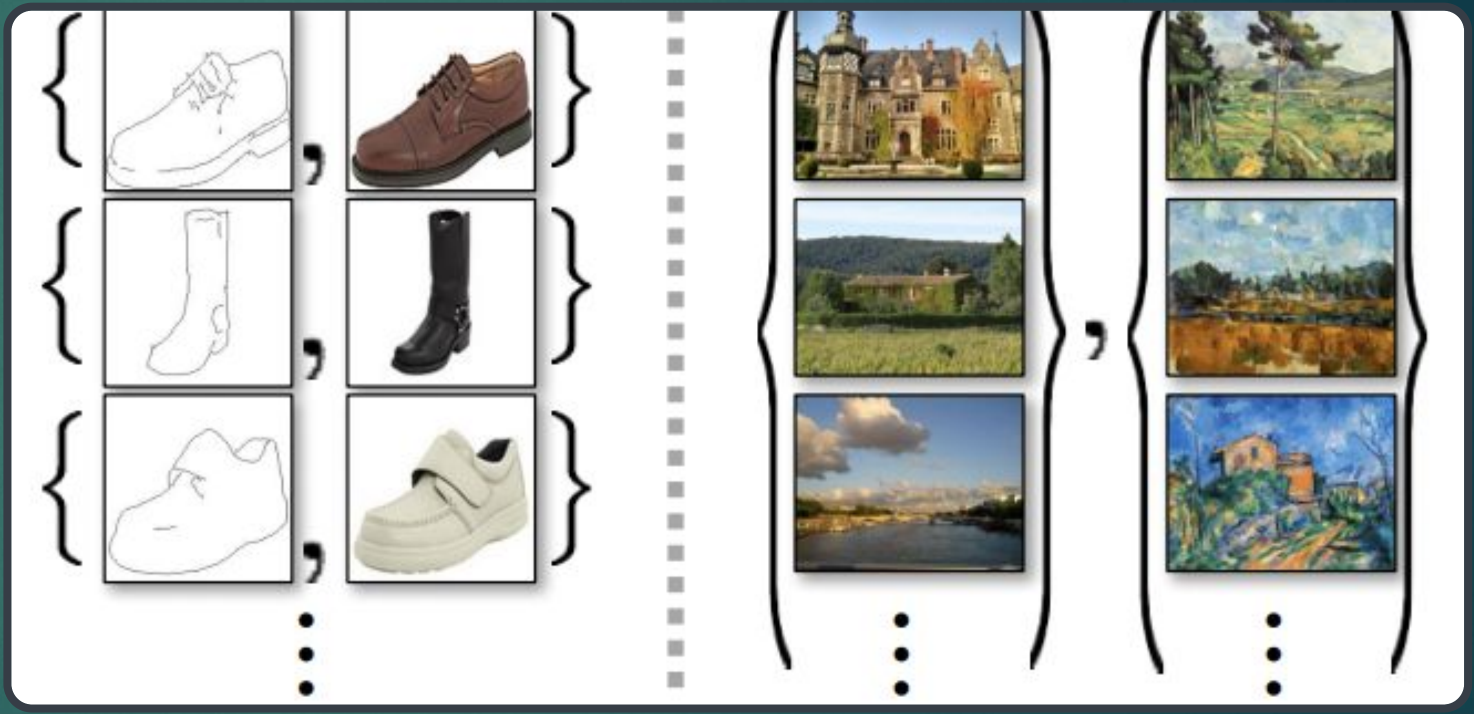
Data Preparation Occurs at Two Steps in Pipeline

- Pre-CycleGAN Phase
  - Generate large number of images with smaller dimensions
  - Properly scale source images to match target image domain
  - Generate artificial data from original data
    - Image perturbation
    - Heterogeneous data
  - Tiled data
- Post-CycleGAN Phase
  - Blending original and generated images
  - Ground truth data

CycleGAN

CycleGAN (PyTorch):

- Cycle-Consistent Generative Adversarial Networks
- Impressive results in image generation and image editing, and representation learning.
- Cycle-Consistent loss to enforce two directions training.



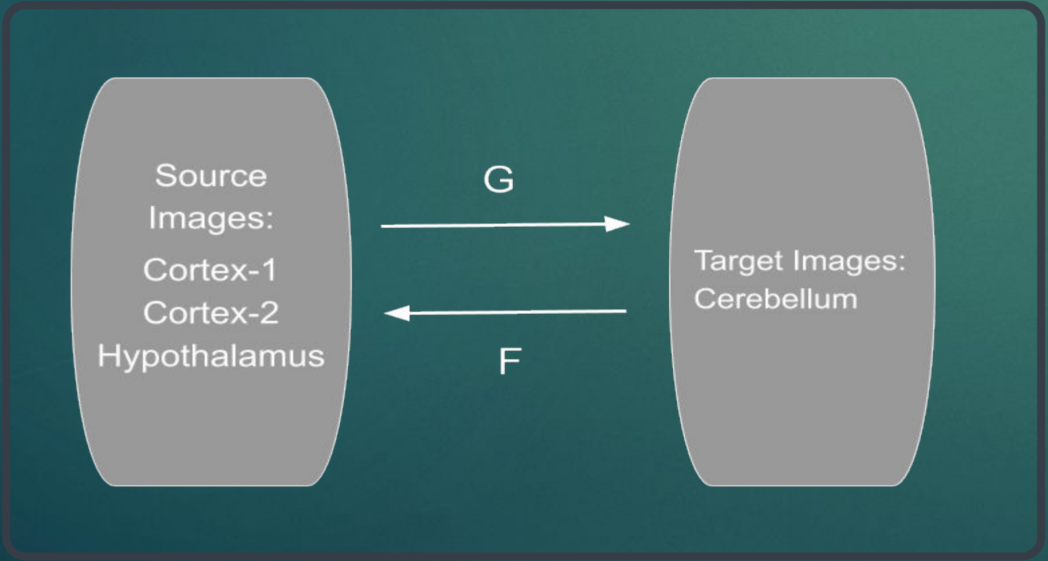
Paired Images

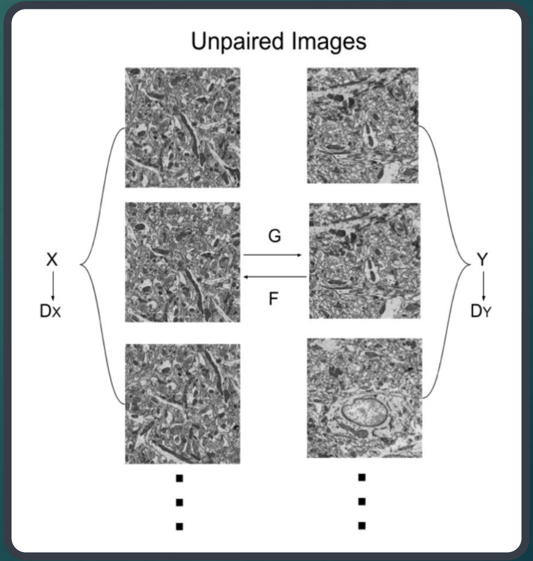
Unpaired Images

CycleGAN(cont.)

CycleGANs algorithm

- Learn a mapping G : X → Y and couple it with an inverse mapping F : Y → X and introduce a cycle consistency loss to push F(G(X)) ≈ X (and vice versa)
- Loss function is a measure of reproducibility of model output
- We can regain input image based on output image

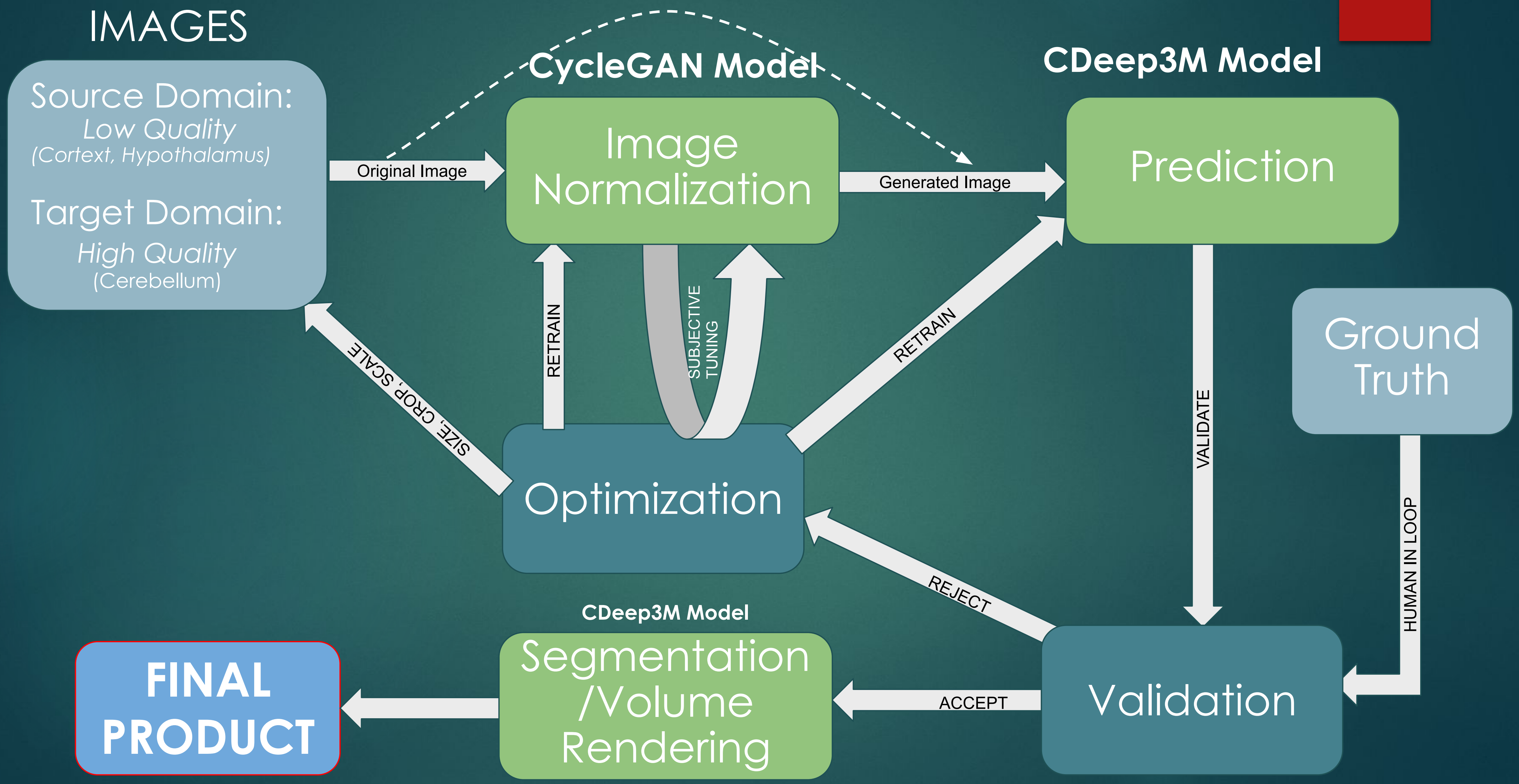




Hypothesis

- CycleGAN will normalize our original images, therefore improving the upstream image segmentation model
    - Image tiling with overlap will improve the poor neural net performance on the edges of tiles
- Retraining the CDeep3M model with normalized images will improve the image segmentation quality

Solution Architecture


IMAGES
Source Domain:
Low Quality
(Cortext, Hypothalamus)
Target Domain:
High Quality
(Cerebellum)
CycleGAN Model
Image Normalization
CDeep3M Model
Prediction
Original Image
Generated Image
RETRAIN
SUBJECTIVE TUNING
RETRAIN
SIZE, CROP, SCALE
Optimization
VALIDATE
Ground Truth
HUMAN IN LOOP
REJECT
CDeep3M Model
Segmentation /Volume Rendering
ACCEPT
Validation
FINAL PRODUCT

CycleGAN Model

Generate Normalized Images

Step 1:

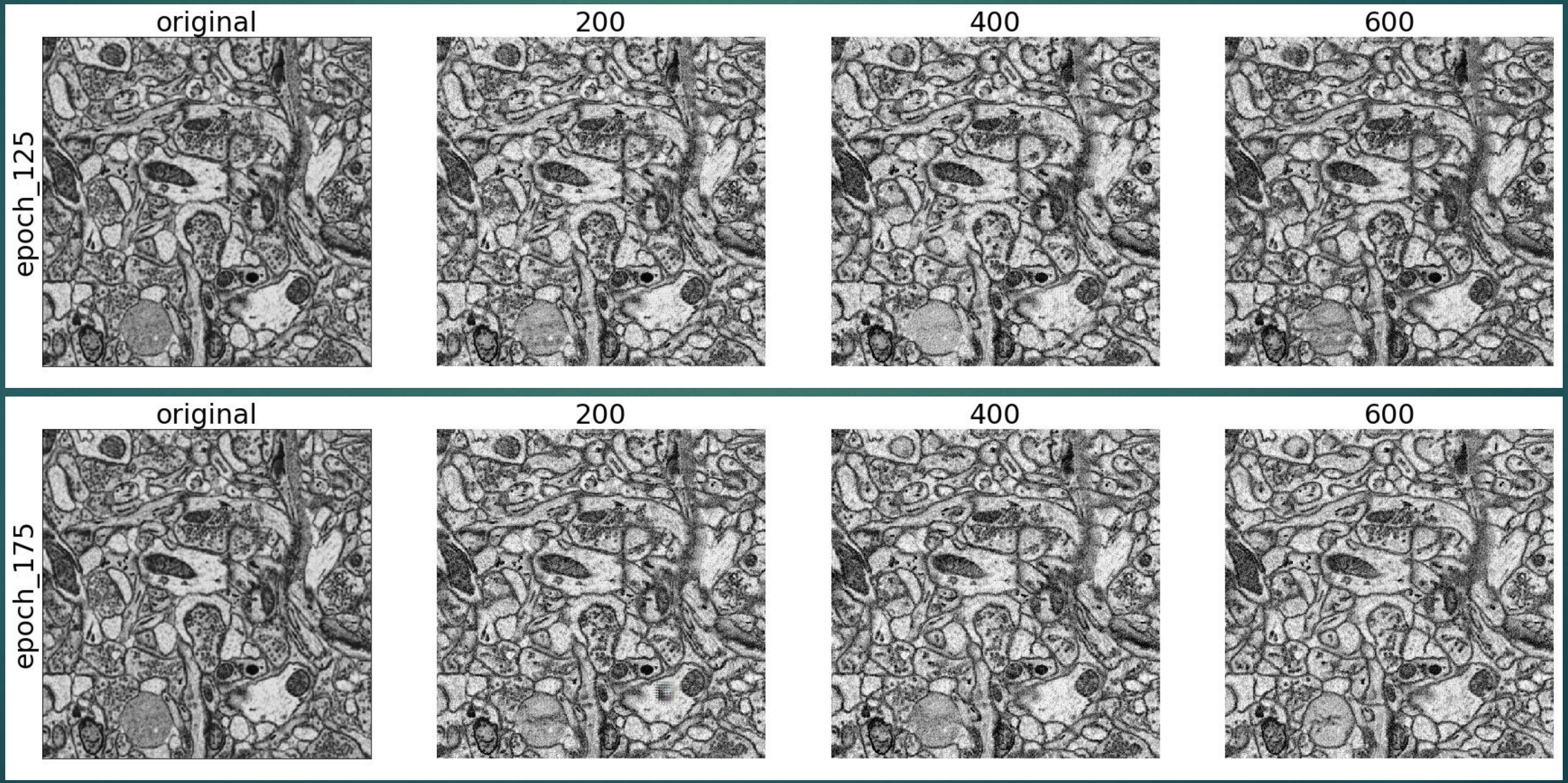
- Creating different datasets
  - Quantity
  - Scaling
  - Heterogeneous data
- Training models (learning rate, epoch)

Different volumes of images with different scaling factors, learning rates and epoch values

Model Exploration - Training Size

Source Domain: Cortex_1 dataset

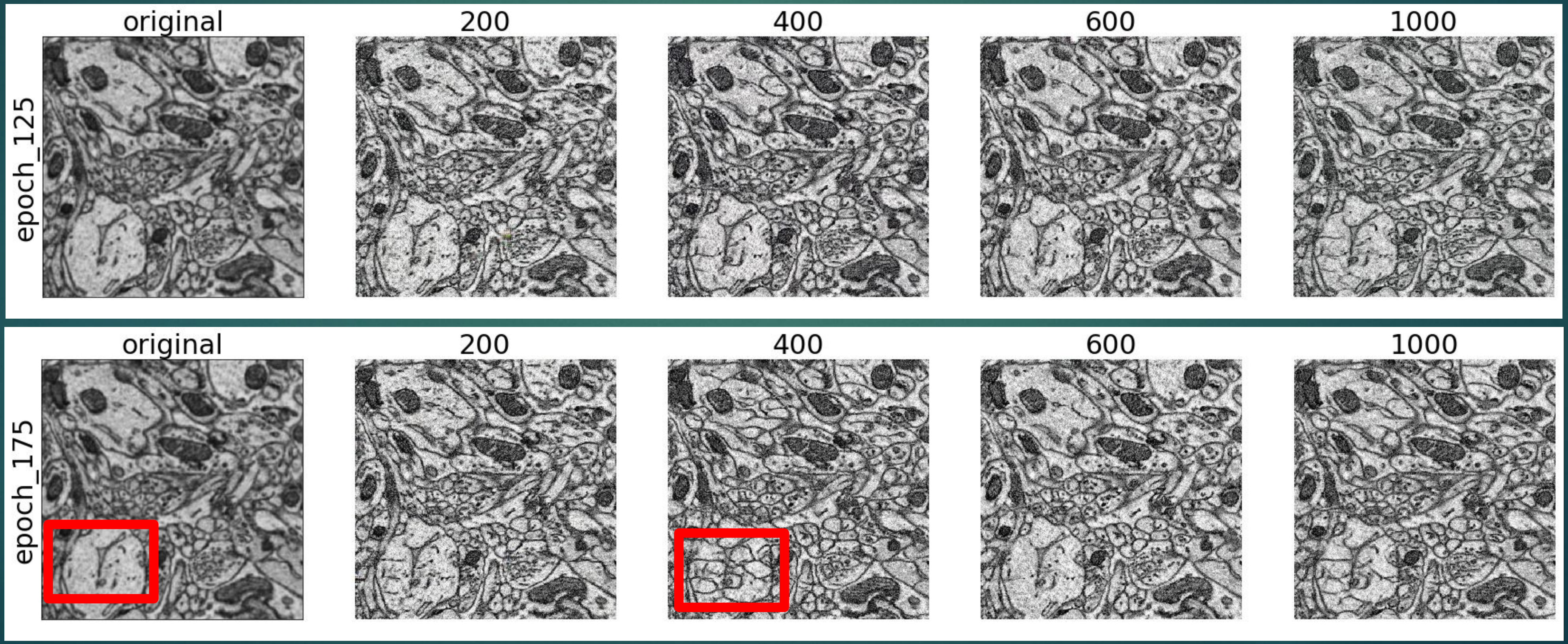
Target Domain: Cerebellum dataset



Model Exploration - Training Size (cont.)

Source Domain: Cortex_1 dataset

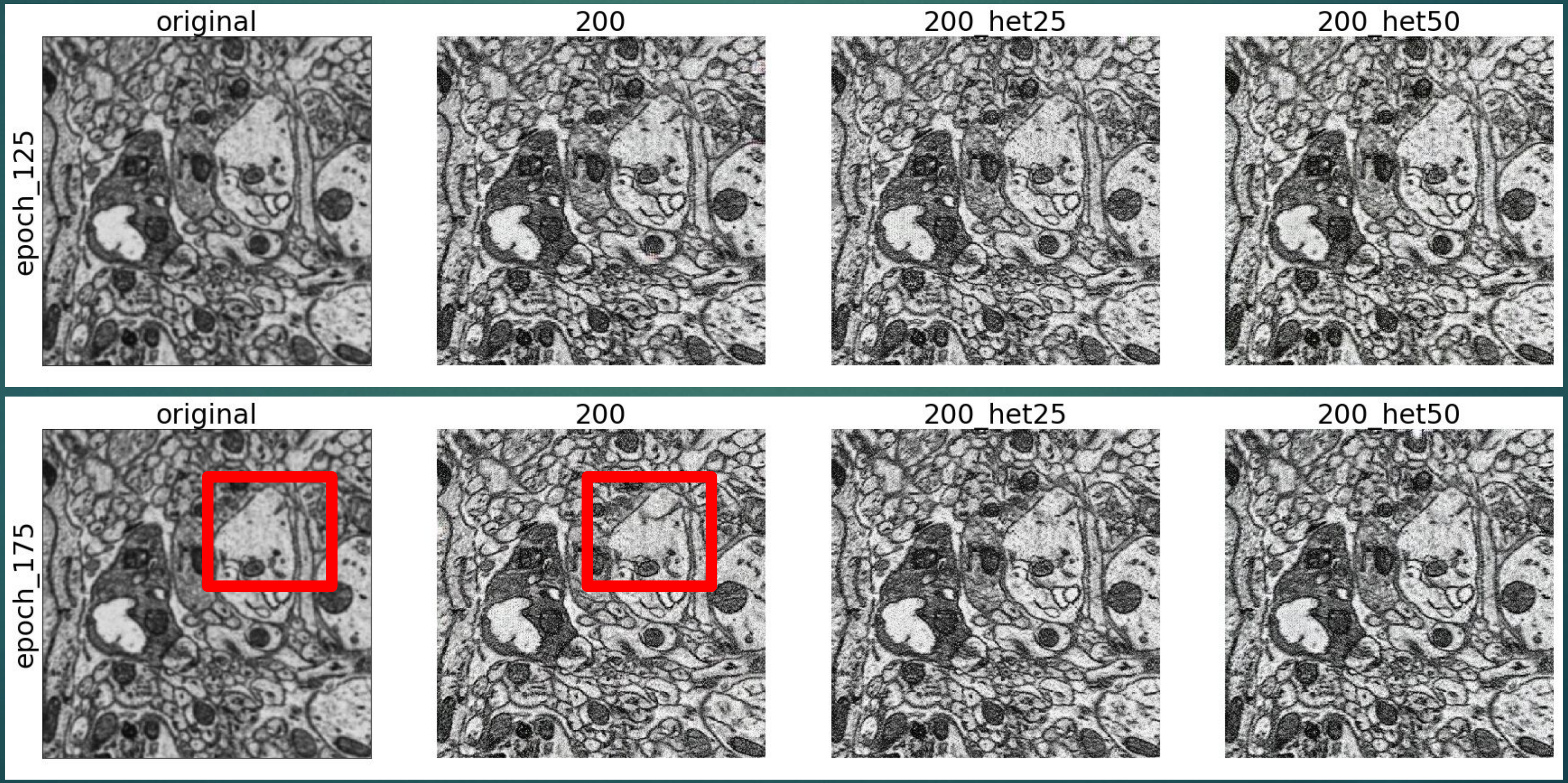
Target Domain: Cerebellum dataset



Model Exploration - Heterogeneous Data

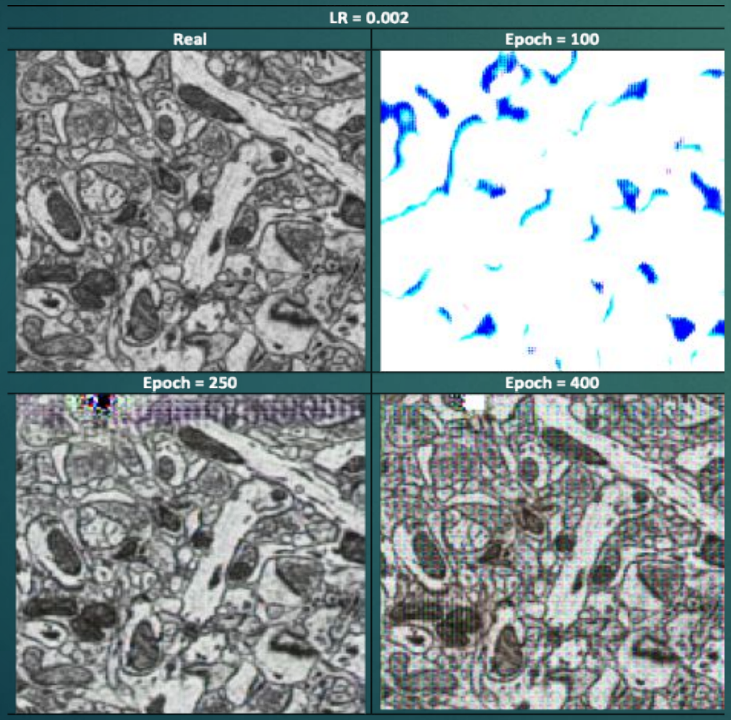
Source Domain: Mixture of hypothalamus data and altered cerebellum dataset

Target Domain: Cerebellum dataset



Model Optimization - Learning Rate

- LR from $2.0x10^{-8}$ to $2.0x10^{-2}$ at the step of 10 times growth have been compared
- LR $2.0x10^{-2}$ does not have coverage at all
- The result of epoch 250 is the best for rates of $2.0x10^{-3}$, $2.0x10^{-4}$, and $2.0x10^{-5}$

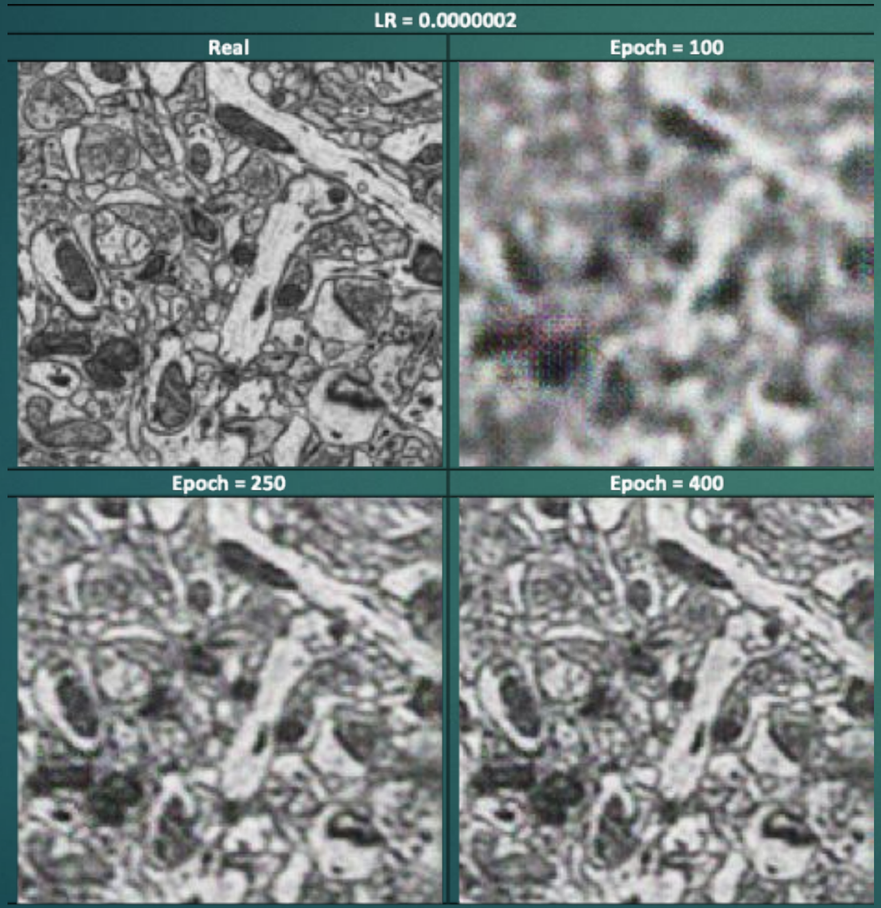


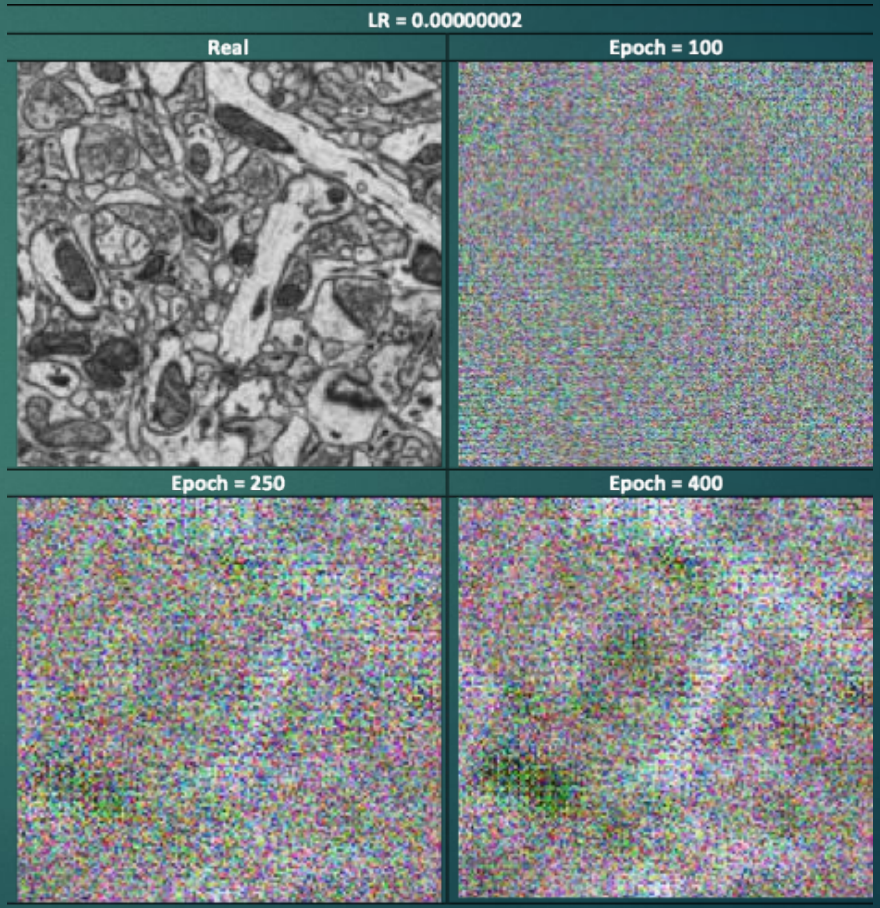
Model Optimization - Learning Rate (cont.)

- For LR $2.0x10^{-6}$, $2.0x10^{-7}$, and $2.0x10^{-8}$, the result of 400 epochs is the best.
- LR $2.0x10^{-6}$ has the best results with 400 epochs among all scenarios
- (confirmed by domain expert)

Model Optimization - Learning Rate (cont.)

Lowest LR (2.0×10^{-8}) with 400 epochs does not have coverage.

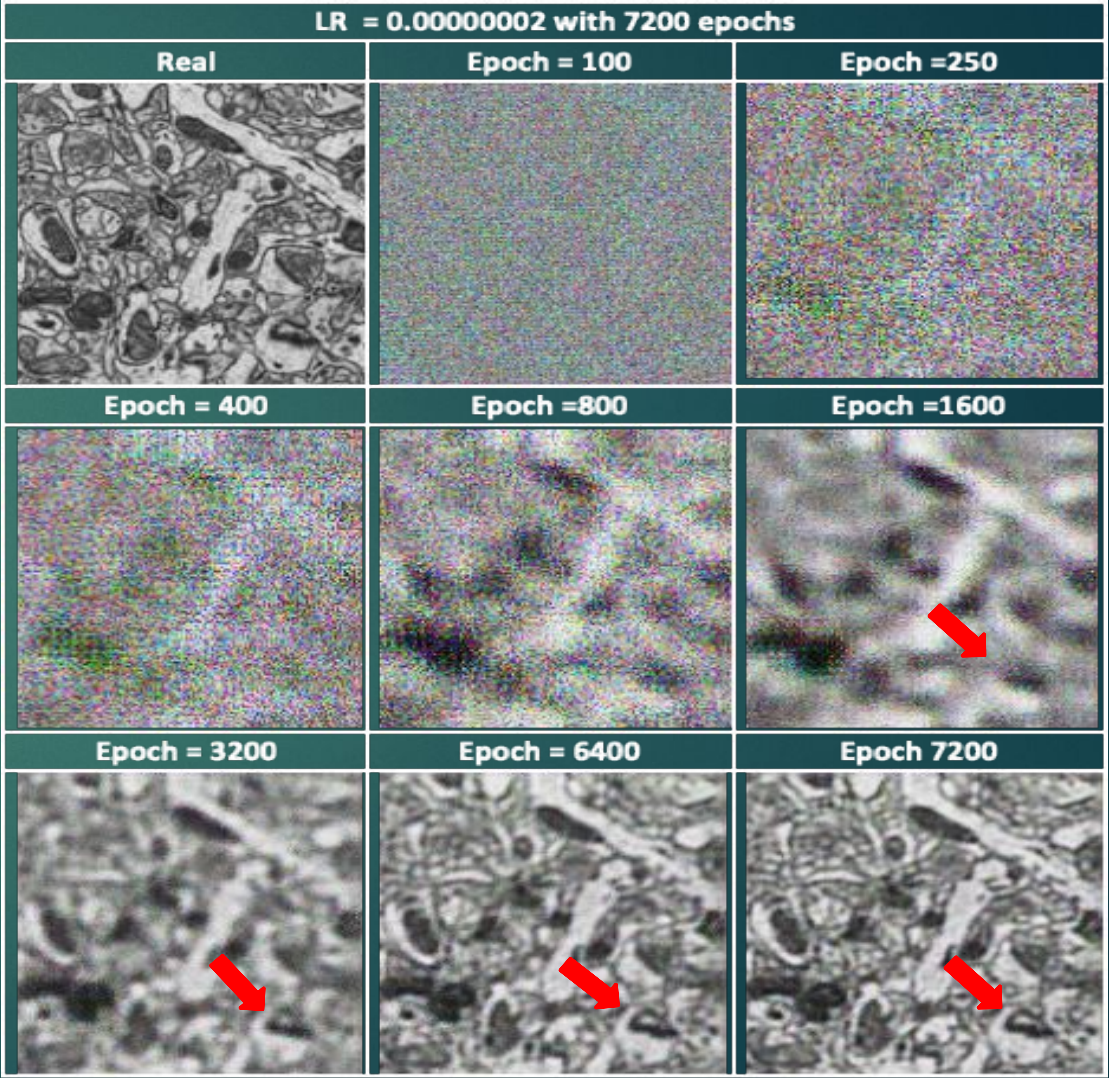




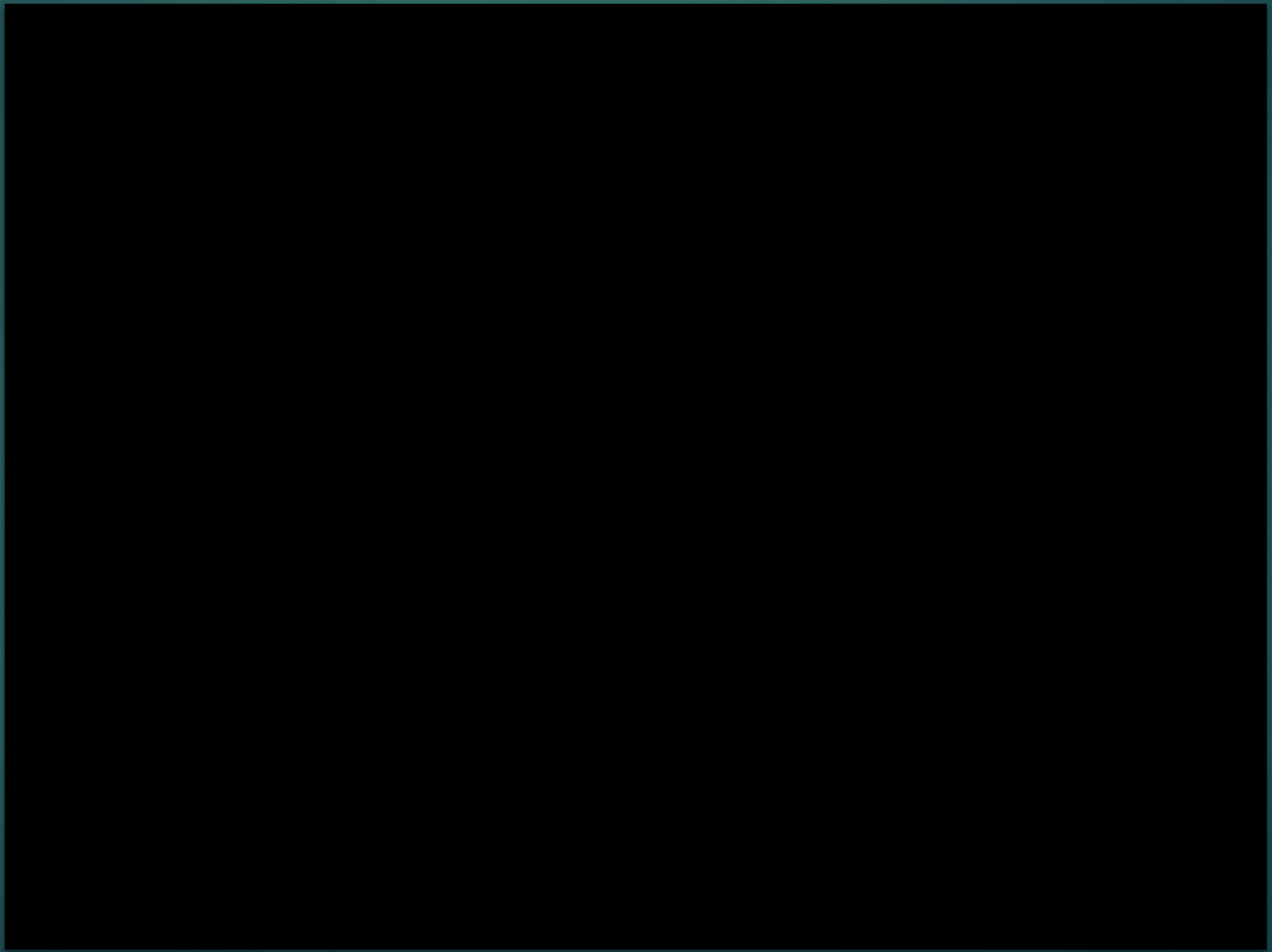
Model Performance - Learning Rate

LR = 2.0×10^{-8}

- Increasing the number of epochs from 250 to 400 improved the quality of the generated image.
- Training continued over larger number of epochs to investigate the performance over time.
- Epoch values of 100, 250, 400, 800, 1600, 3200, 6400 and 7200 have been tested.
- The tests had a run time of over **80 hours**, while using **multiple GPUs** with batch processing.



Model Performance



CycleGAN Model (cont.)

Step 2:

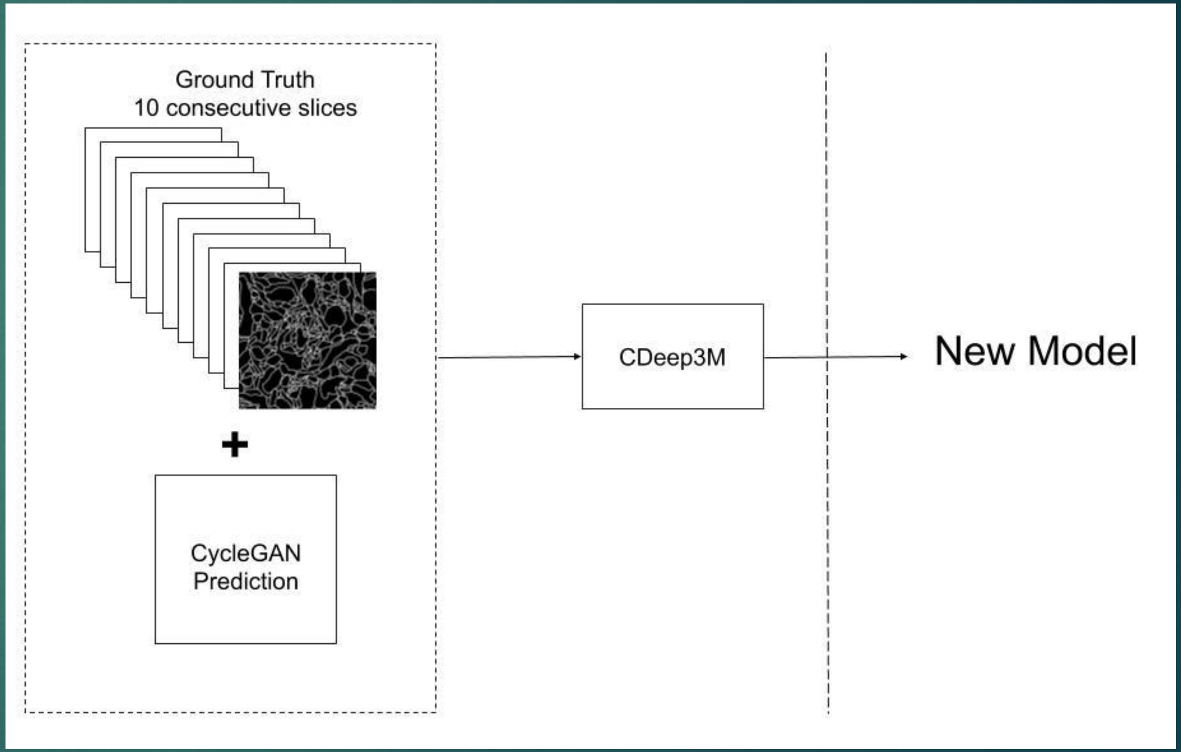
- Pre-processing:
  - Taking the input images and tiling them
- CycleGAN output:
  - Modified Tiled Images
- Post-processing:
  - Fusing the tiled images to get the full sized normalized image

Enhancing edge quality

CDeep3M Model

Step 3:

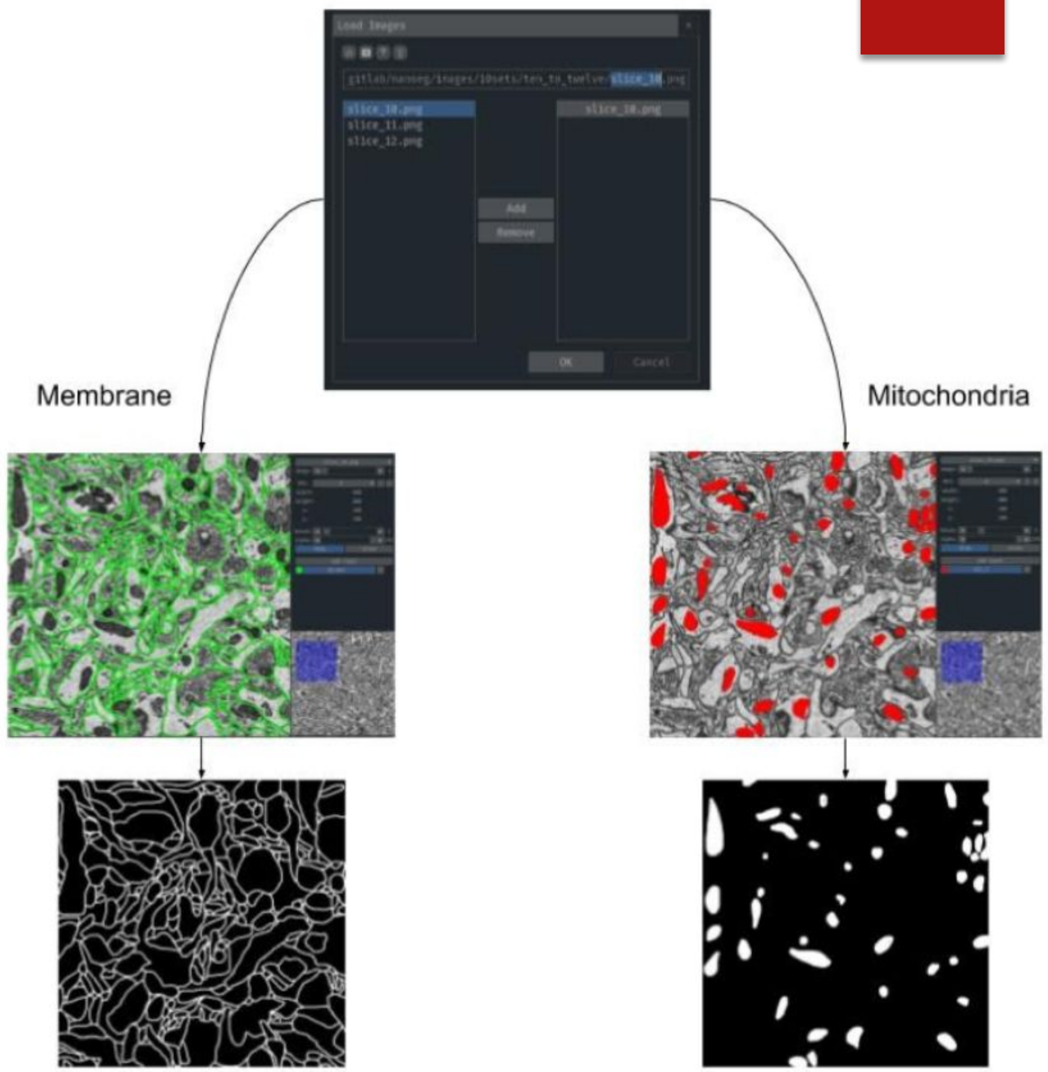
- Creating blended datasets with CycleGAN generated images and original images
- Retraining the model with ground truth images and the blended images



Segmentation Tool

Creating Ground Truth

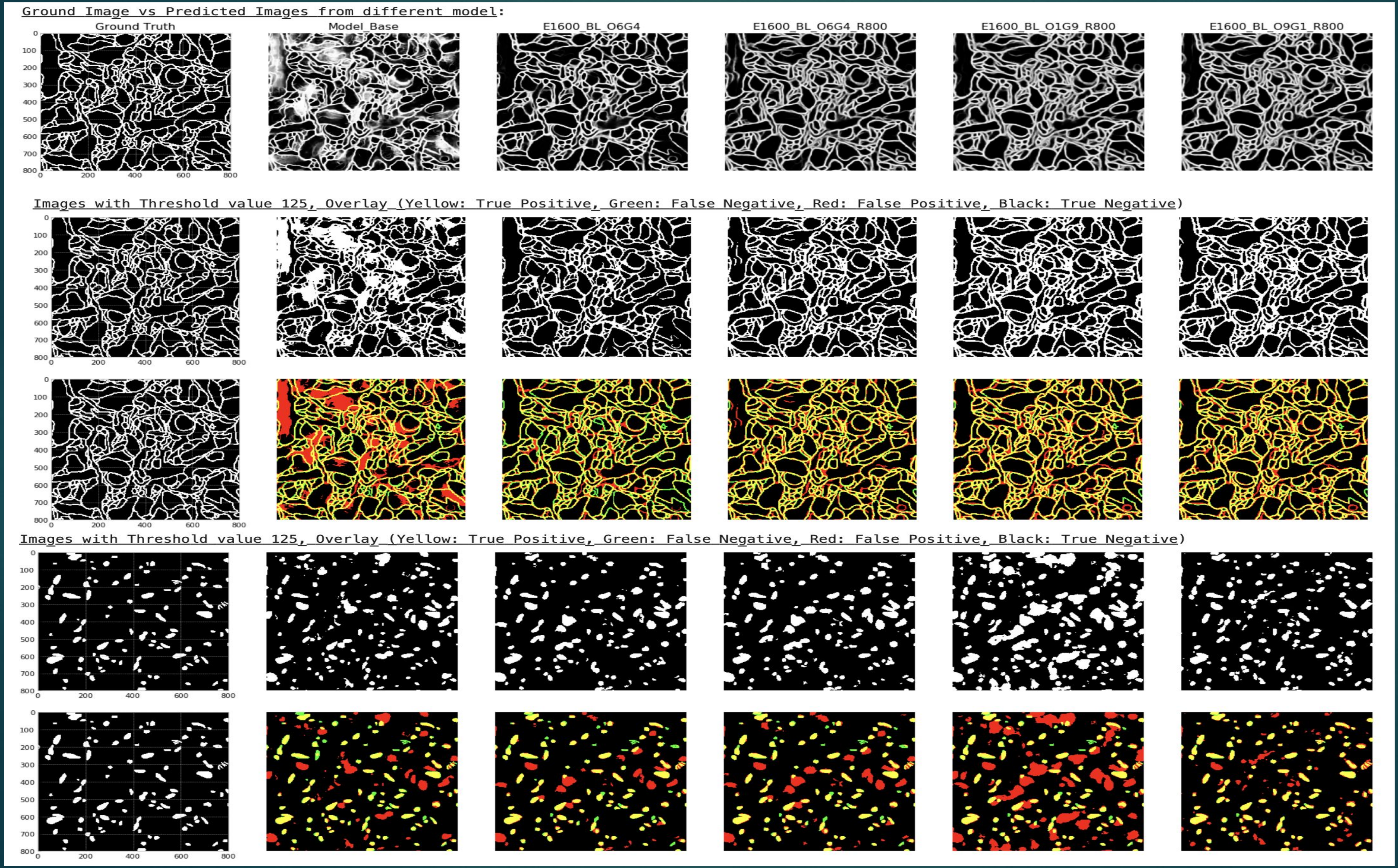
- 10 Consecutive slices (membranes and mitochondria)
- Same ROI
- 800x800 pixel crop



Validation and Performance Evaluatation

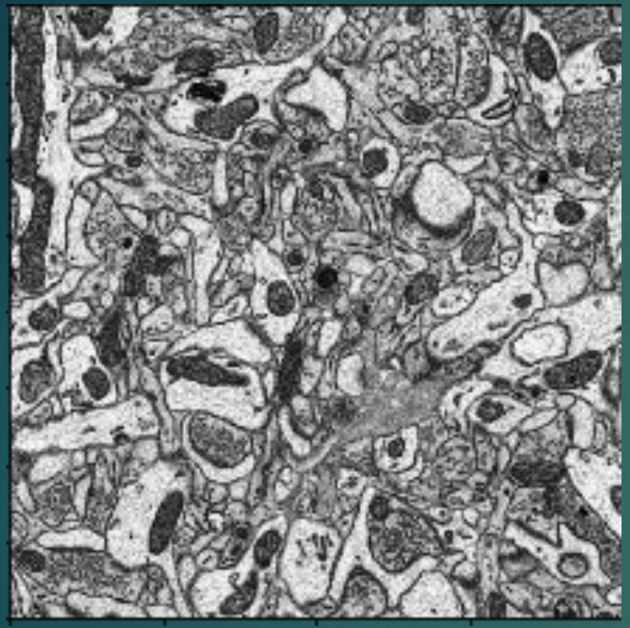
- Subjective Measurement
  - Qualitative
  - Image overlay
- Quantitative Measurement
  - Accuracy, precision, recall, F-Score, F-beta
  - Why is F-Beta important?
- Model Selection from Evaluation of 40+ Models
  - Both measurements done with different pixel threshold values 100,125, 150,...,250
  - With best F1-Score while keeping F-Beta in consideration too

Qualitative Evaluation

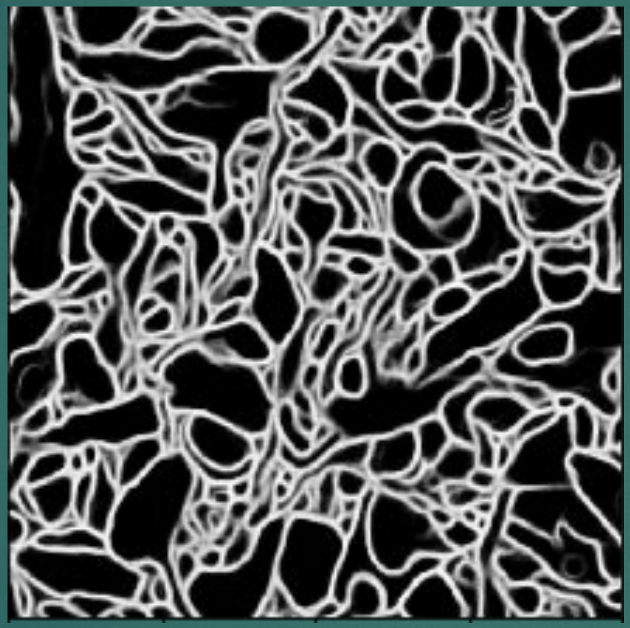


Yellow	True Positive
Green	False Negative
Red	False Positive
Black	True Negative

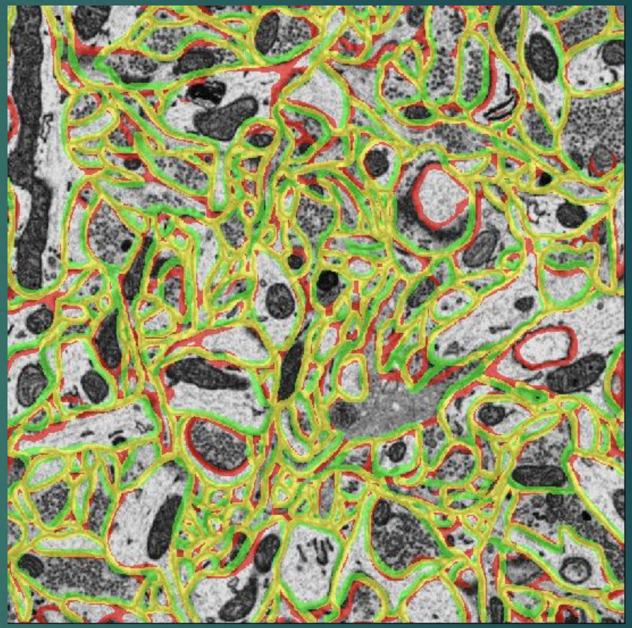
Image Overlay



Original Image



Predicted Membranes



Yellow	True Positive
Green	False Negative
Red	False Positive

Quantitative Evaluation

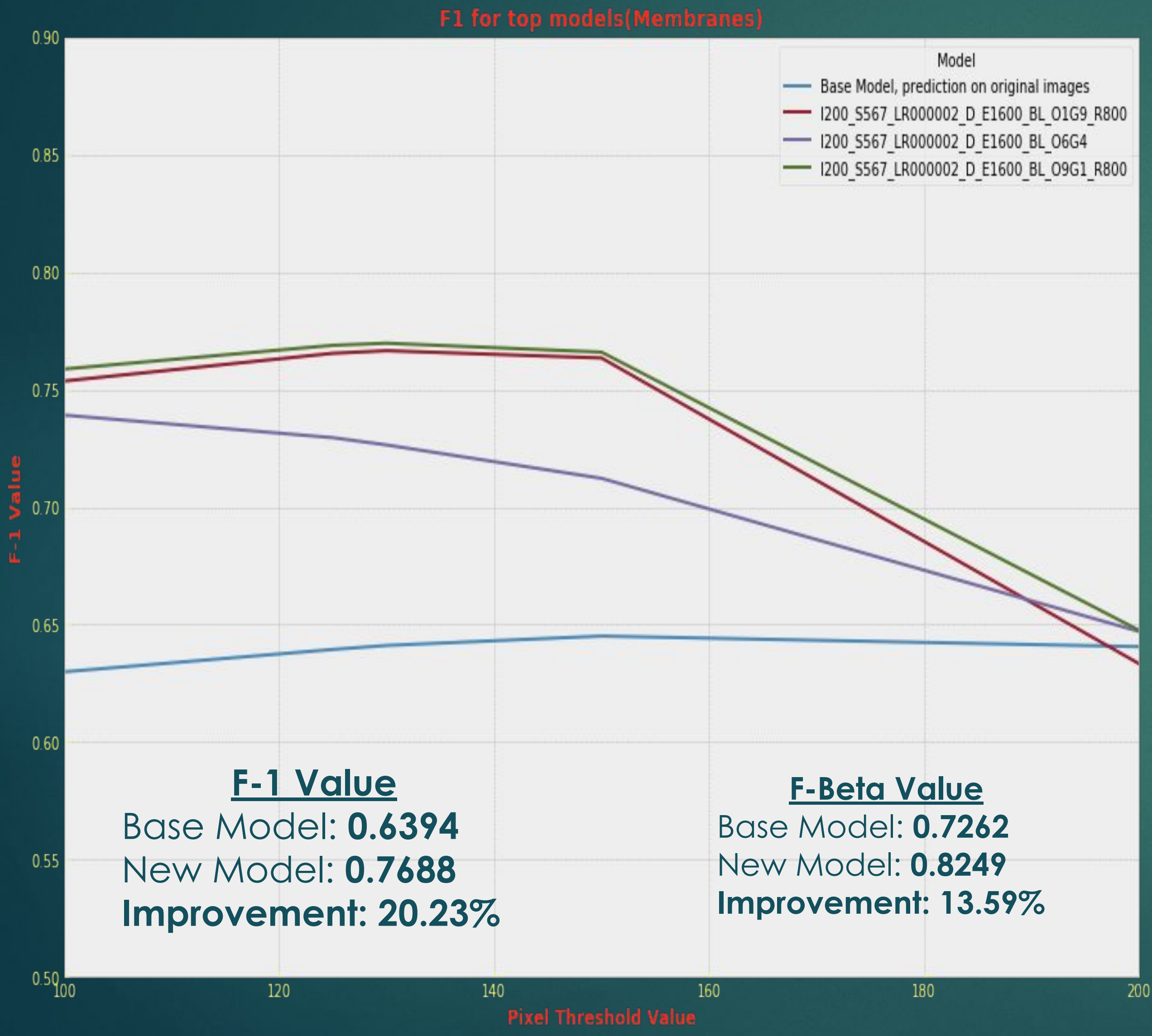
Membranes

Model	Threshold	Images	Scale	Epoch	Denoise	Learning Rate	Retrained	Blended	Original(%)	Generated(%)	Precision	Recall	Specificity	Accuracy	F1	F_Beta
I200_S567_LR000002_D_E1600_BL_O1G9_R800: Re-trained Model (+800 iterations) with generated & groundtruth data, prediction on blended(10% Original 90% Generated) images, and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	Yes	Yes	10	90	0.6823	0.8716	0.8355	0.8459	0.7654	0.8258
I200_S567_LR000002_D_E1600_BL_O9G1_R800: Re-trained Model (+800 iterations) with generated & groundtruth data, prediction on blended(90% Original 10% Generated) images, and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	Yes	Yes	90	10	0.6907	0.867	0.8426	0.8497	0.7688	0.8249
I200_S567_LR000002_D_E1600_BL_O6G4_R800: Re-trained Model (+800 iterations) with generated & groundtruth data, prediction on blended(60% Original 40% Generated) images, and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	Yes	Yes	60	40	0.7074	0.8394	0.8593	0.8536	0.7678	0.8092
I200_S567_LR000002_D_E1600_BL_O6G4: Base Model , prediction on blended(60% Original 40% Generated) images and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	No	Yes	60	40	0.7229	0.7363	0.8856	0.8426	0.7295	0.7336
Base Model, prediction on original images	125		Single	200	No	0.0002	No	No	100	0	0.5332	0.7985	0.7167	0.7403	0.6394	0.7262

Mitochondria

Model	Threshold	Images	Scale	Epoch	Denoise	Learning Rate	Retrained	Blended	Original(%)	Generated(%)	Precision	Recall	Specificity	Accuracy	F1	F_Beta
I200_S567_LR000002_D_E1600_BL_O9G1_R800: Re-trained Model (+800 iterations) with generated & groundtruth data, prediction on blended(90% Original 10% Generated) images, and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	Yes	Yes	90	10	0.6413	0.929	0.9487	0.9469	0.7588	0.8525
I200_S567_LR000002_D_E1600_BL_O1G9: Base Model , prediction on blended(10% Original 90% Generated) images and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	No	Yes	10	90	0.5935	0.8117	0.9451	0.9331	0.6857	0.7561
I200_S567_LR000002_D_E1600_BL_O3G7: Base Model , prediction on blended(30% Original 70% Generated) images and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	No	Yes	30	70	0.5404	0.8211	0.931	0.9211	0.6518	0.7438
Base Model, prediction on original images	125		Single	200	No	0.0002	No	No	100	0	0.5279	0.7534	0.9334	0.9172	0.6208	0.6941
I200_S567_LR000002_D_E1600_BL_O9G1: Base Model, prediction on blended(90% Original 10% Generated) images and tuned hyperparameters	125	200	Multi	1600	Yes	0.000002	No	Yes	90	10	0.3297	0.8618	0.8269	0.83	0.4769	0.6515

Quantitative Evaluation (cont.)

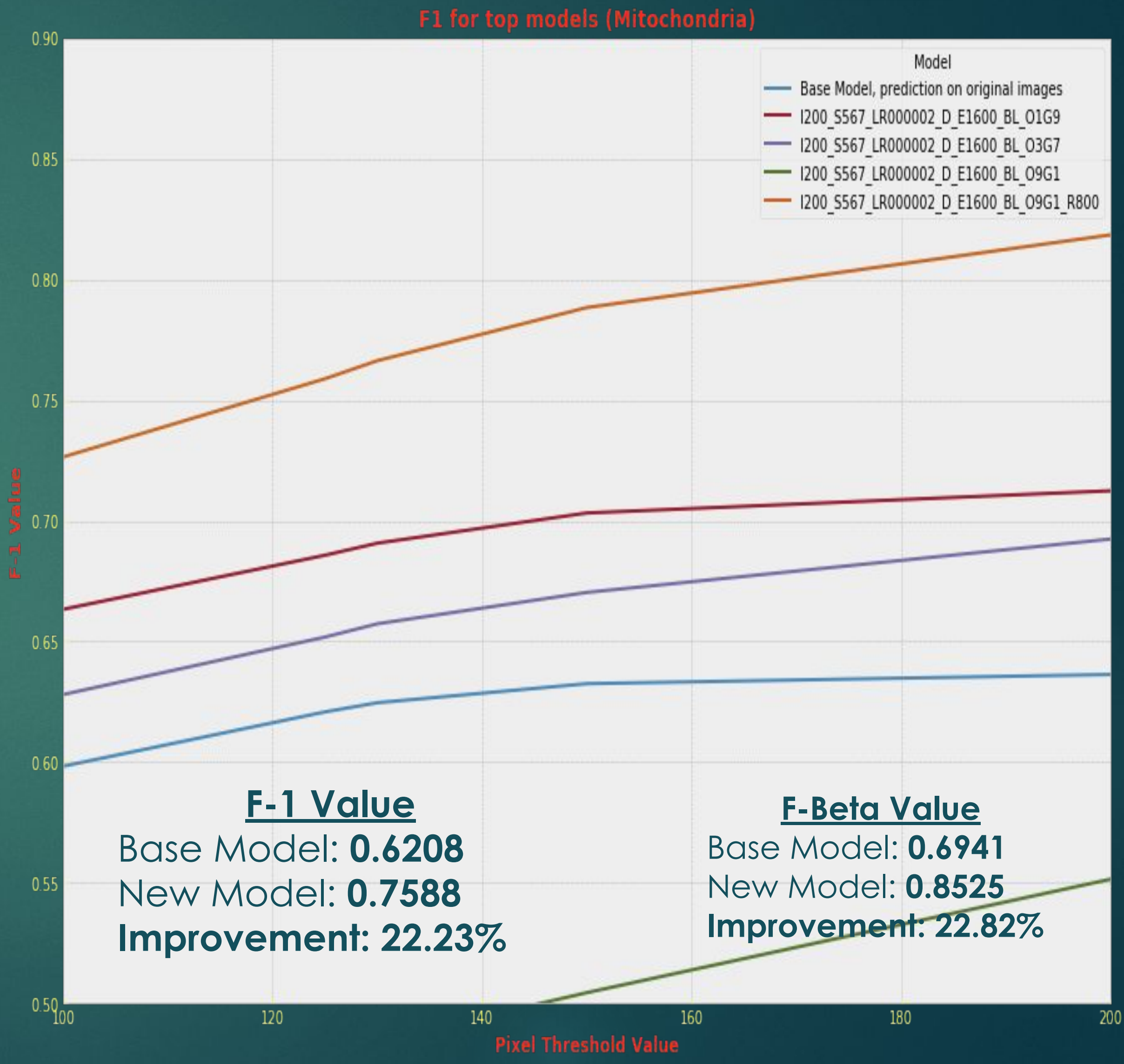


F-1 Value

Base Model: **0.6394**
New Model: **0.7688**
Improvement: 20.23%

F-Beta Value

Base Model: **0.7262**
New Model: **0.8249**
Improvement: 13.59%



F-1 Value

Base Model: **0.6208**
New Model: **0.7588**
Improvement: 22.23%

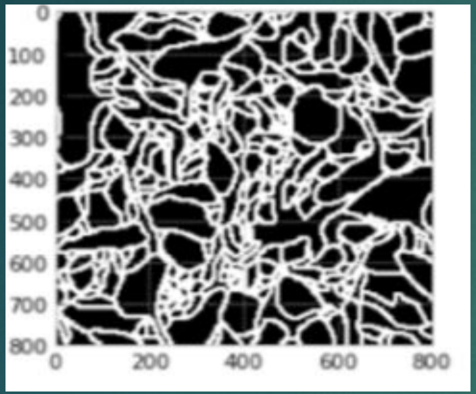
F-Beta Value

Base Model: **0.6941**
New Model: **0.8525**
Improvement: 22.82%

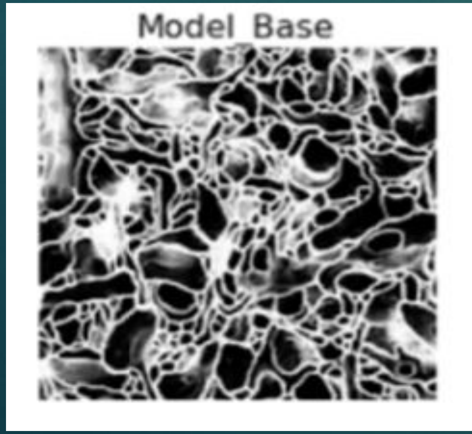
Model Selection

Membrane

Ground Truth



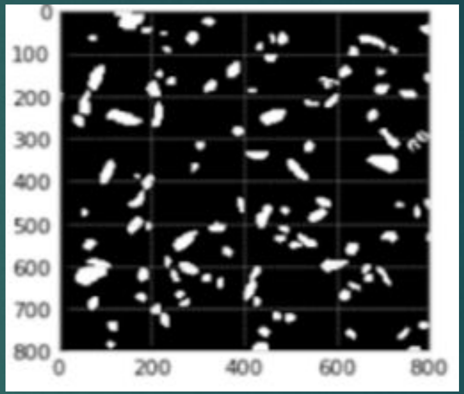
Base Model Output



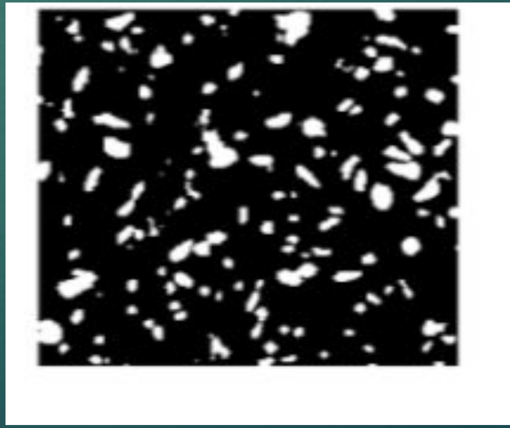
New Model Output

Mitochondria

Ground Truth



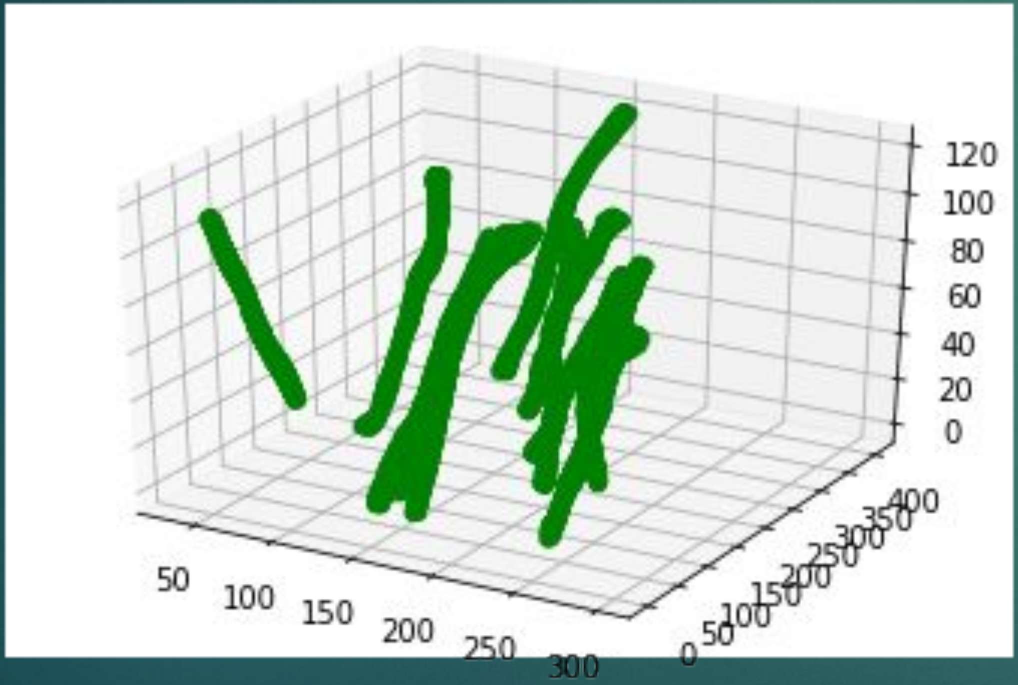
Base Model Output



New Model Output

Volume Rendering

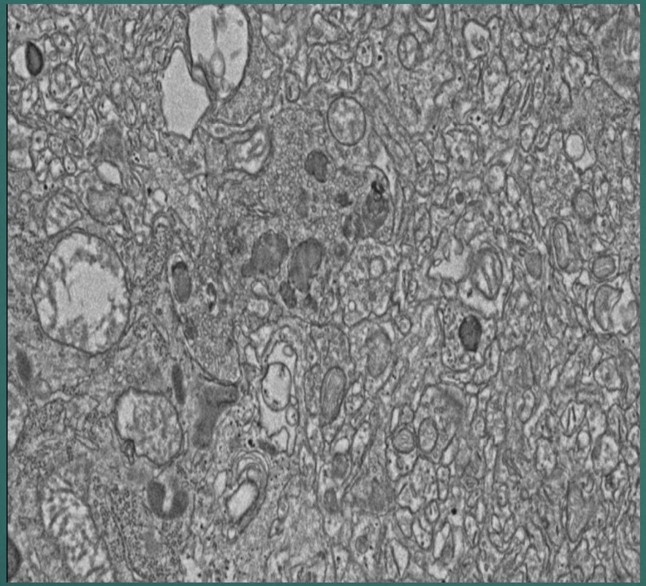
Before



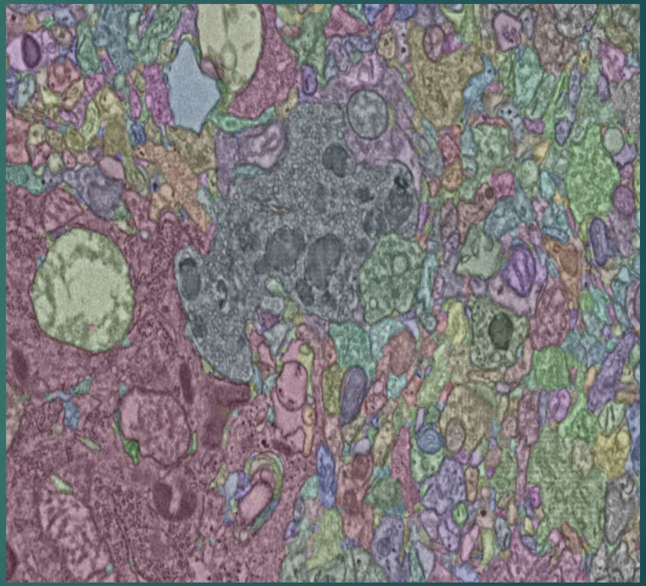
After

Future Work

Techniques and findings applied to current projects can be applied to other domains (e.g. human tissue samples)



Before



After

Conclusion

Accomplishments/Findings

- Measured Improvement in Membrane Detection
  - Over 20% improvement in F1 score
  - Approx 14% improvement in F-beta score
- Measured Improvement in Mitochondria Detection
  - Over 22.2% improvement in F1 score
  - Approx 22.8% improvement in F-beta score
- Retraining leads to thicker Membrane Predictions
  - Cleaner predictions with good boundary
- Interactive Volume Rendering
  - End user has more control

Acknowledgements

- Advisors
  - Matthew Madany
  - Dr. Mark Ellisman
  - Steve Peltier
- Dr. Ilkay Altintas
- UCSD MAS DSE Program and Staff
- Friends and Family
- Various pets and random smurfs

Questions?