Deep Learning-Based Nuclei Segmentation of Cleared Brain Tissue

Pooya Khorrami[1], Kevin Brady[1], Mark Hernandez[1], Lars Gjesteby[1], Sara Nicole Burke[2], Damon Lamb[2], Matthew A. Melton[2], Kevin Otto[2], Laura J. Brattain[1]

[1]MIT Lincoln Laboratory

[2]University of Florida

IEEE HPEC 2019

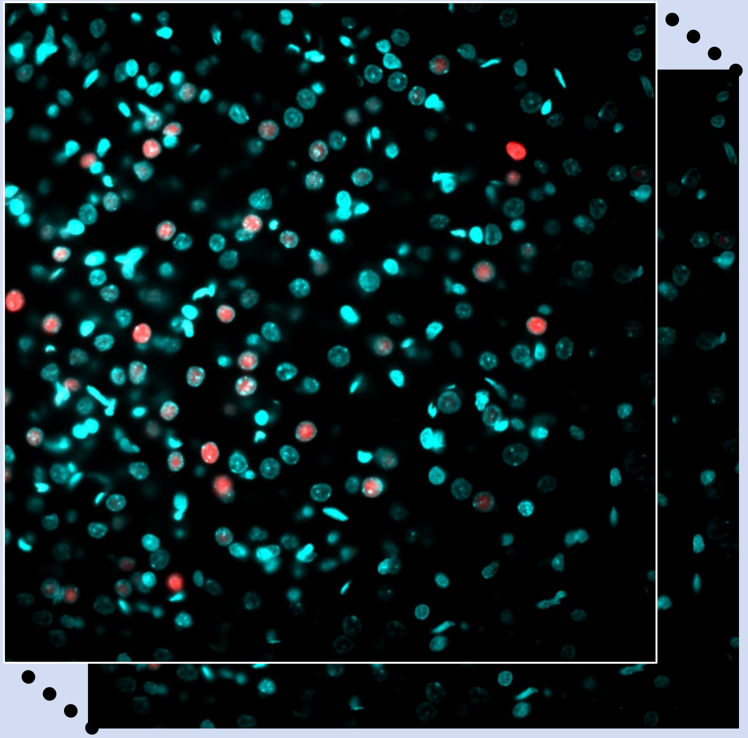
26 September 2019

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

This work was also sponsored by the Defense Advanced Research Projects Agency (DARPA) BTO under the auspices of Dr. Douglas Weber and Dr. Tristan McClure-Begley through the DARPA Contracts Management Office Grant No. HR0011-17-2-0019.

Program Goal, Challenges, and Proposed Approach

Light-sheet Microscopy Images



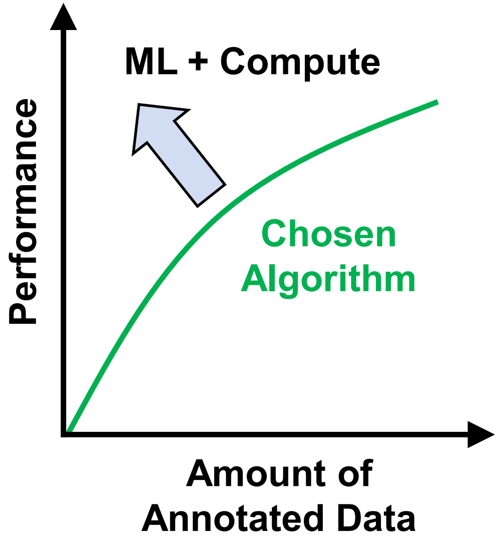
Goal is to map neuron connectivity

Applications

- 3D Nuclei Segmentation
- Centroid Detection
- Axon Tracing
- ⋮

Challenges: Difficult to process large-scale, densely populated data automatically with high accuracy

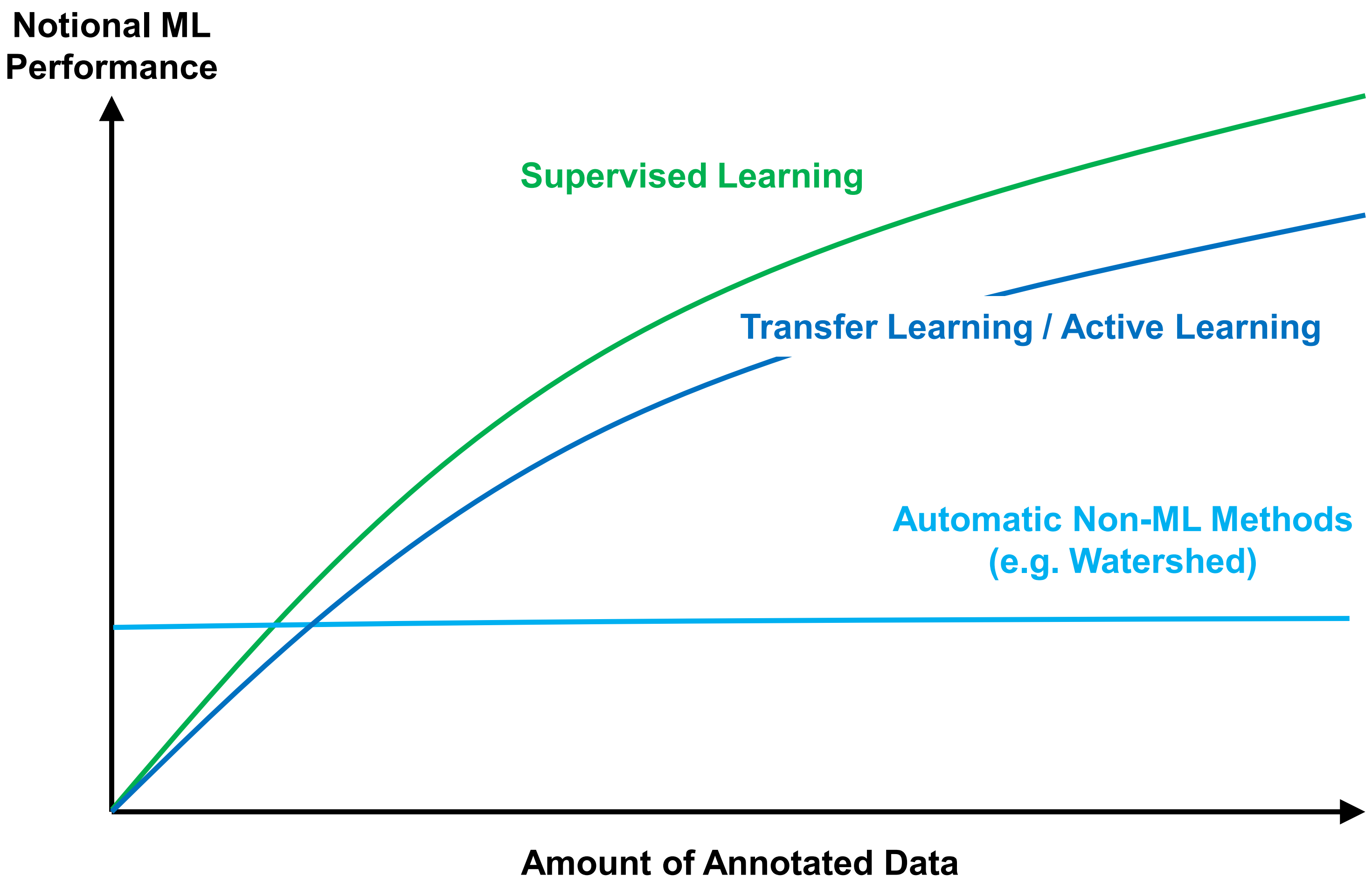
Evaluation



Propose Approach: Develop HPC-based ML algorithms

ML: Machine Learning

Concept Overview



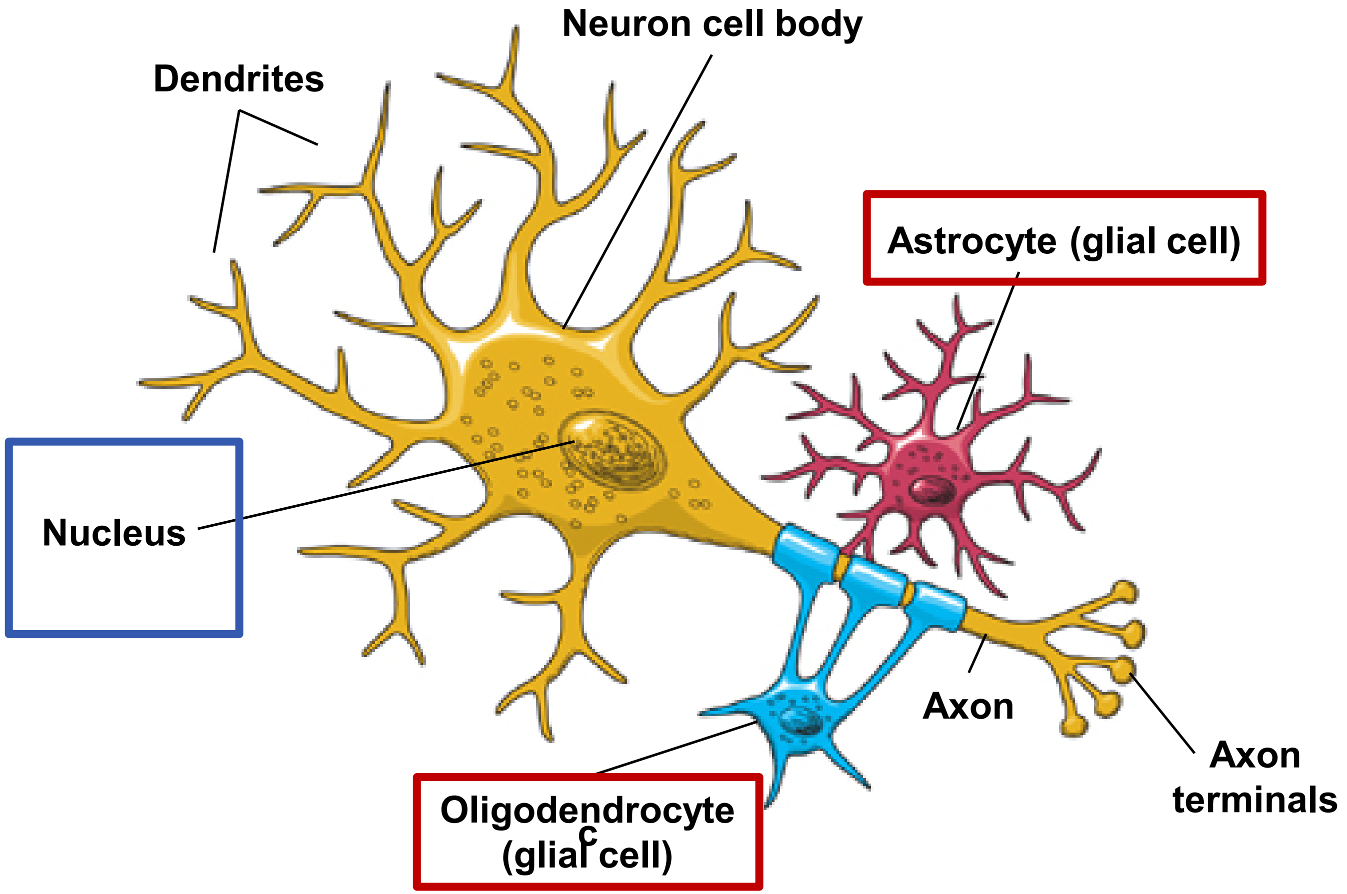
Available Approaches

- Traditional Approach
  - Collect and annotate more data
  - Challenge: time+resource intensive
- Lincoln Approach
  - Leverage automatic and/or semi-automatic algorithms to obtain comparable accuracy with less annotation

ML: Machine Learning

Targets of Interest

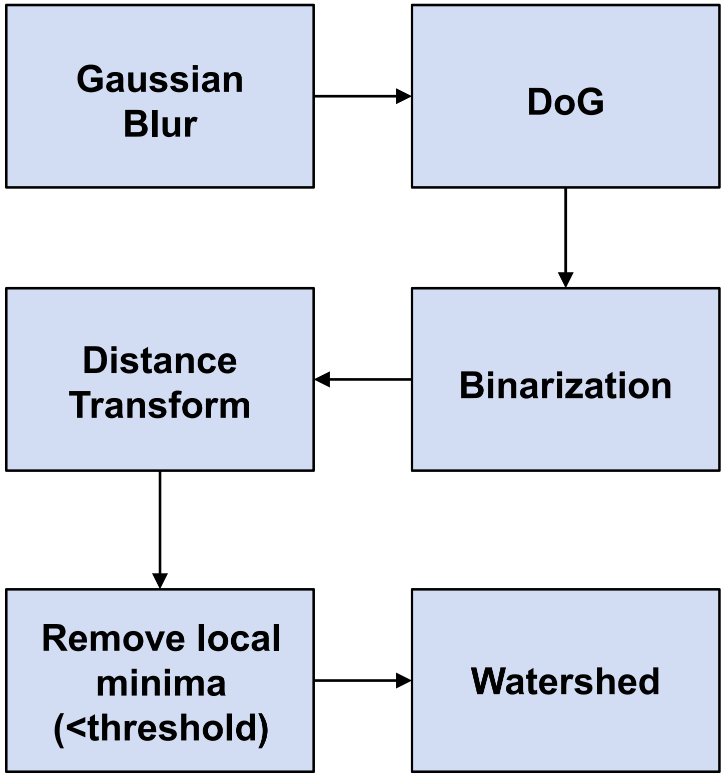
- Neurons (nuclei)
  - Receive stimuli
  - Conduct action potentials
- Glial cells
  - Supporting functions
  - Hold neurons close together



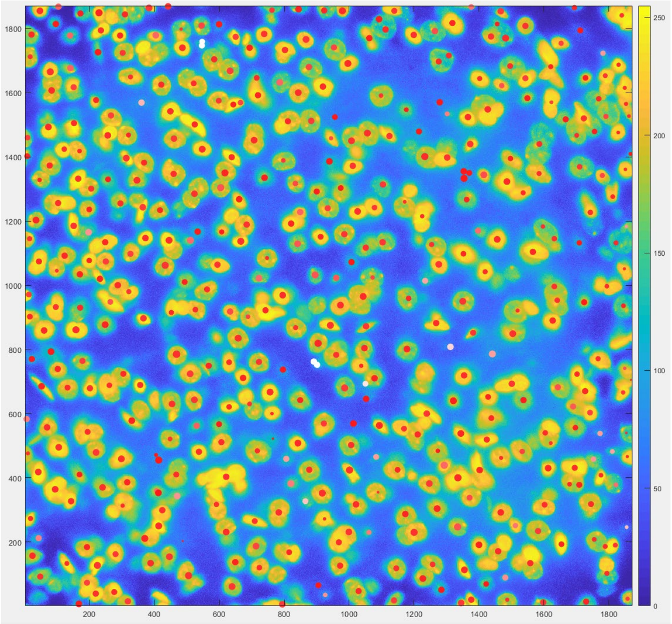
Source: NIH Institute of Neurological Disorders and Stroke

Automatic Approach
Conventional Image Processing on LLSC

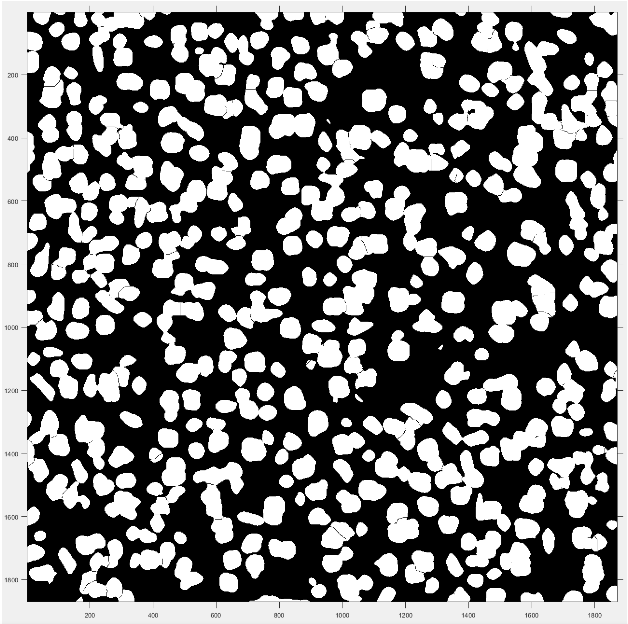
Difference of Gaussian (DoG), Thresholding, and Watershed Segmentation



Centroid Detections



Binary Mask



Strength
- No need for annotation
- Results can be fed to ML algorithm as weakly annotated data

Weakness
- Bulk blob detection method does not distinguish between different cell types (i.e. neurons and glia)

LLSC: Lincoln Laboratory Supercomputing Center

Proposed Technical Approach
Reduce costly human annotation

Existing model

U-Net Model on nuclei segmentation

Transfer Learning

Our datasets

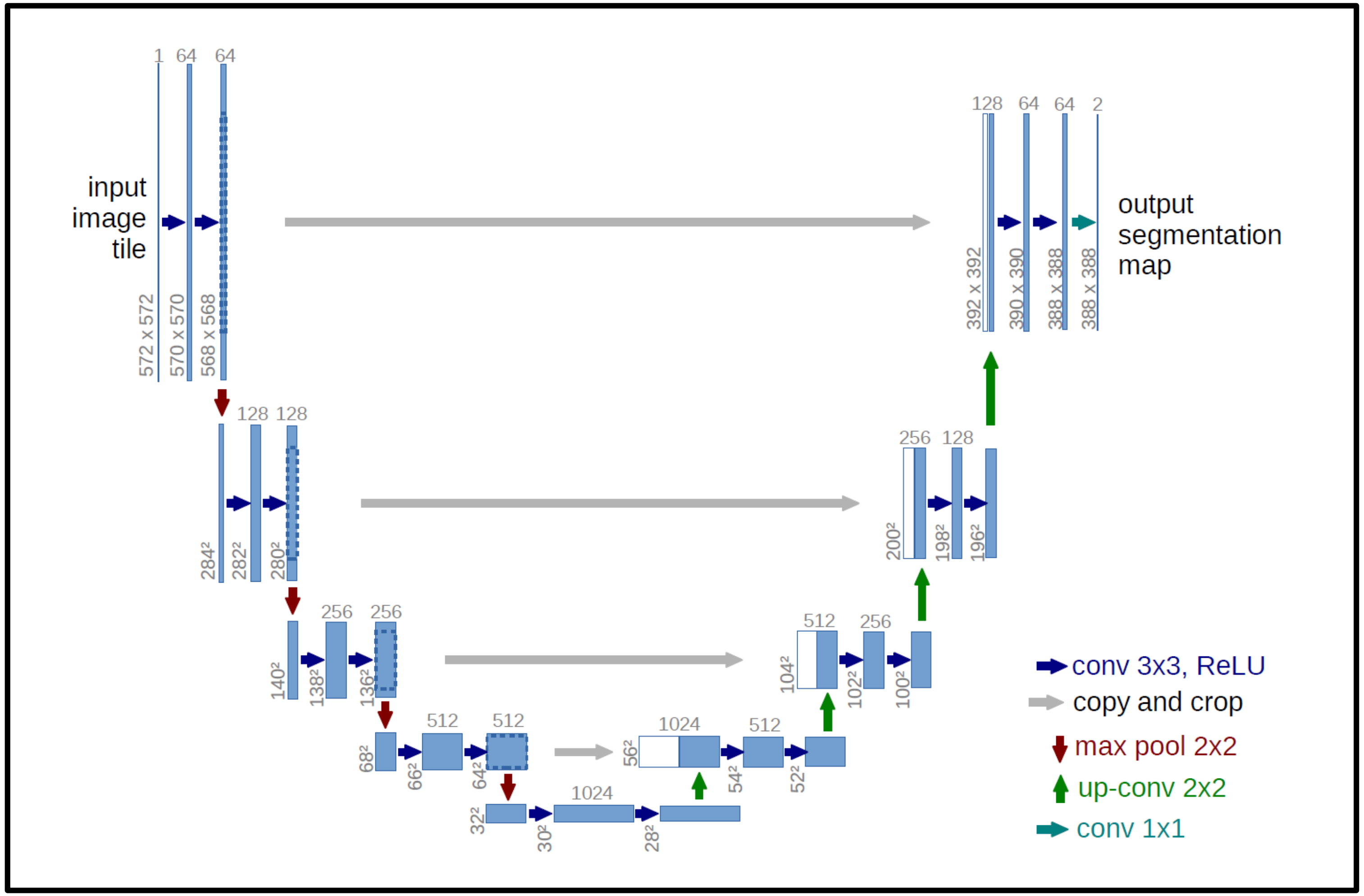
- **Large amount of unlabeled nuclei data**
- **Small amount of labeled nuclei data**

Results

- **Qualitative results**
- **Quantitative results**

Transfer Learning Approach
U-Net

- **U-Net [1] is a popular type of "fully-convolutional" neural network**
- **It is comprised of a contracting path (left) and an expanding path (right)**
  - **Expanding path uses information from contracting path (via "copy and crop" operations)**
- **Contains 23 convolutional layers in total**



[1] - O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

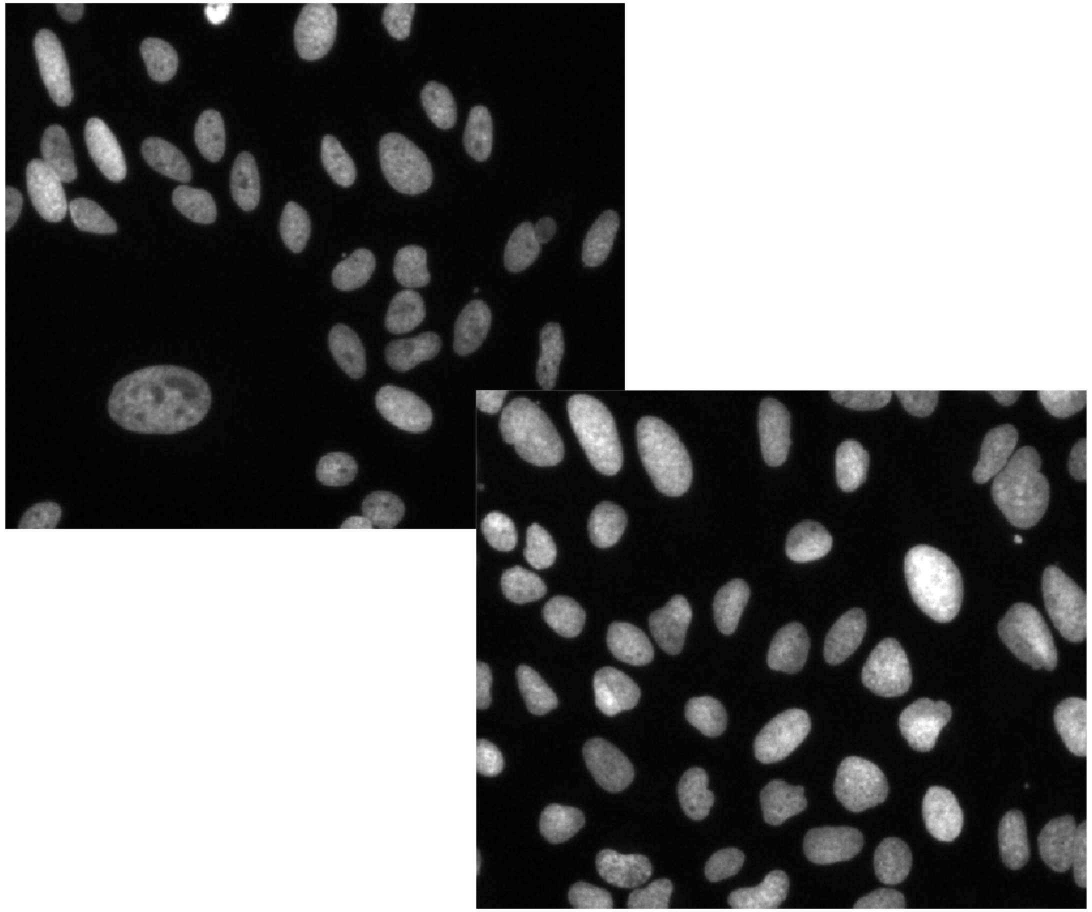
Proxy Dataset

Used nuclei segmentation dataset prepared by the Broad Institute (BBBC039) [1, 2]

200 images collected – fluorescence microscopy

Around 23,000 nuclei annotated

Image dimensions: 520 x 626 pixels



[1] - V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation." Nature methods, vol. 9, no. 7, pp. 637–637, 2012.
[2] - https://data.broadinstitute.org/bbbc/BBBC039/

Proposed Technical Approach
Reduce costly human annotation

Existing model

U-Net Model on nuclei segmentation

Transfer Learning →

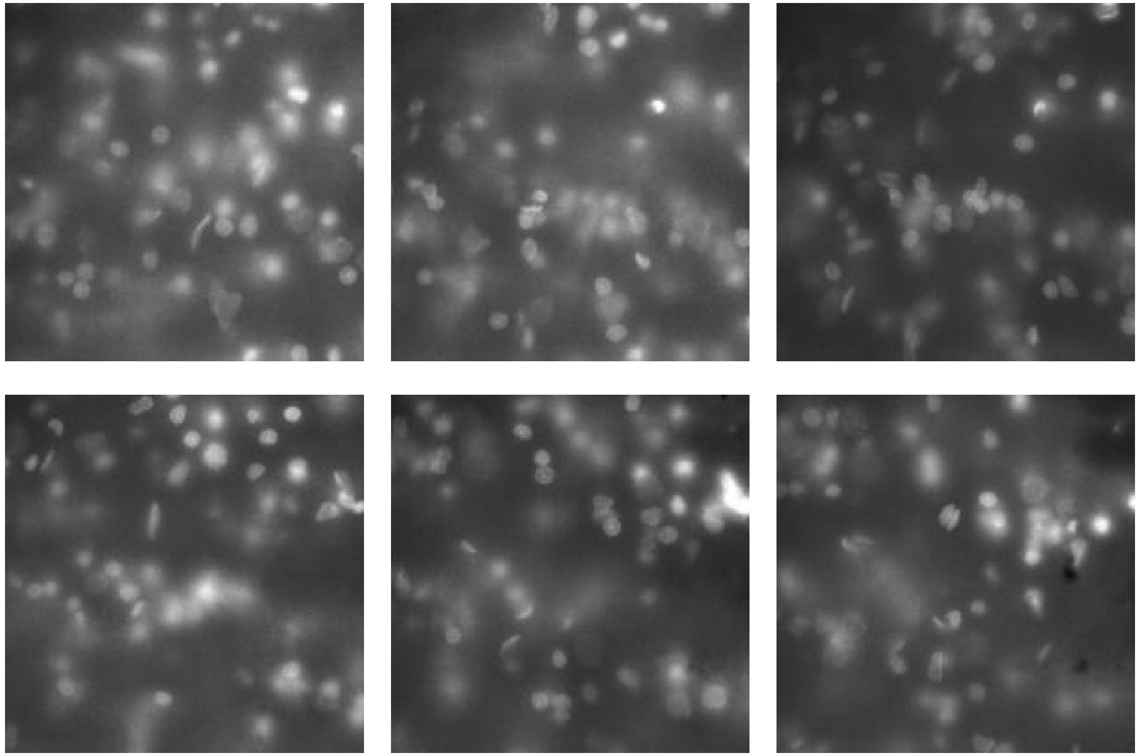
Our datasets
- Large amount of unlabeled nuclei data
- Small amount of labeled nuclei data

→

Results
- Qualitative results
- Quantitative results

Unlabeled Data (DAPI)

- Rat brain tissue sample of the nucleus tractus solitarus region
- Prepared using CLARITY tissue clearance [1]
- Applied a fluorescent DNA stain, diamidino-2-phenylindole (DAPI)
- Image acquired using a light-sheet microscope (ZEISS Lightsheet Z.1)
- Resolution:
  - 1920 x 1920 x 650 voxels
  - 0.2µm x 0.2µm x 1µm



Examples of DAPI slices

[1] - K. Chung and K. Deisseroth, “Clarity for mapping the nervous system,” Nature methods, vol. 10, no. 6, p. 508, 2013.

U-Net Training Details

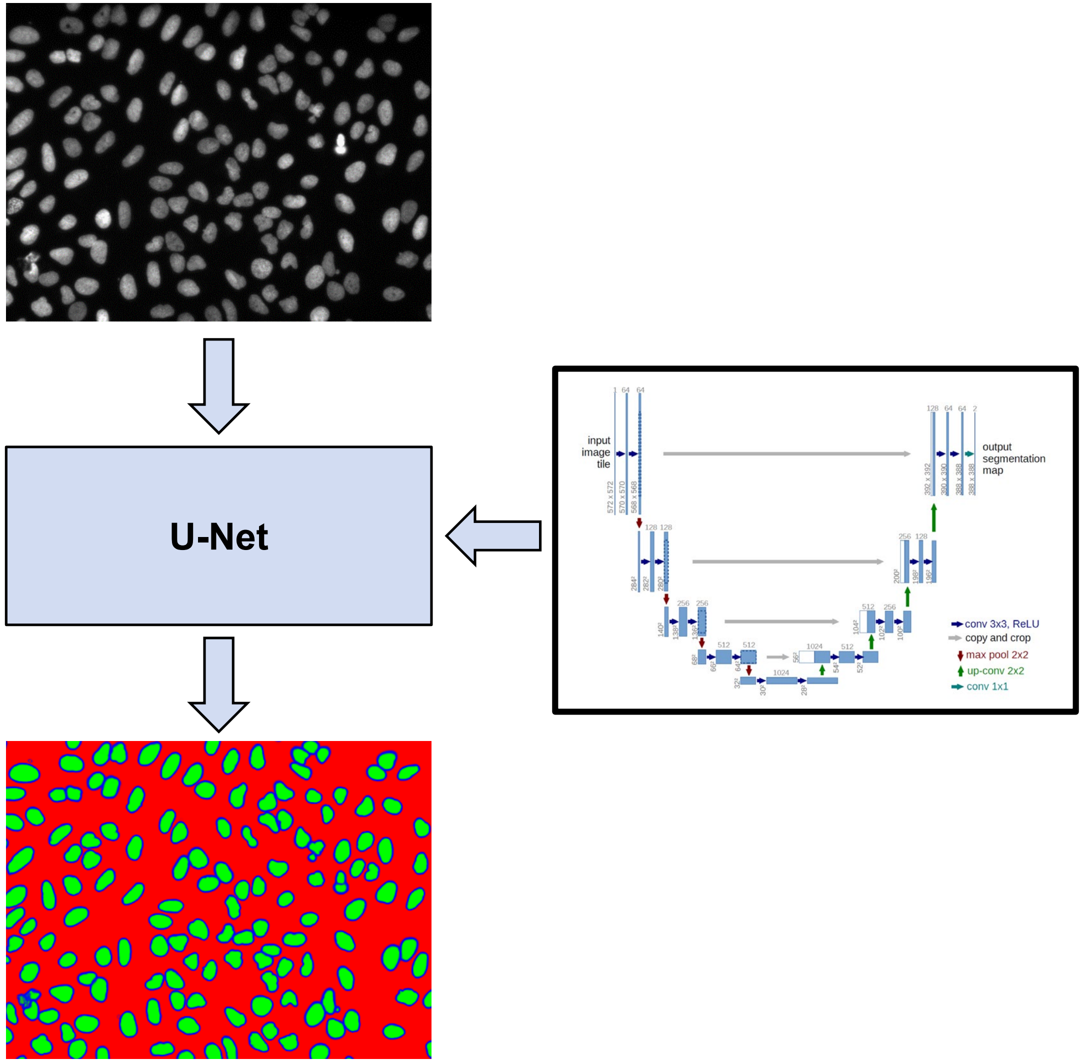
Trained U-Net on BBBC039 dataset using code/library associated with [1]

- Model outputs predictions as being one of three classes:
  1. Background
  2. Nuclei Interior
  3. Nuclei Boundary

Model was trained using Keras with Tensorflow backend on Nvidia Tesla K40

Used MapReduce on LLSC to apply trained model to DAPI slices

- Processed 20 blocks with ~650 slices each
- Watershed Segmentation Approach was also run on LLSC



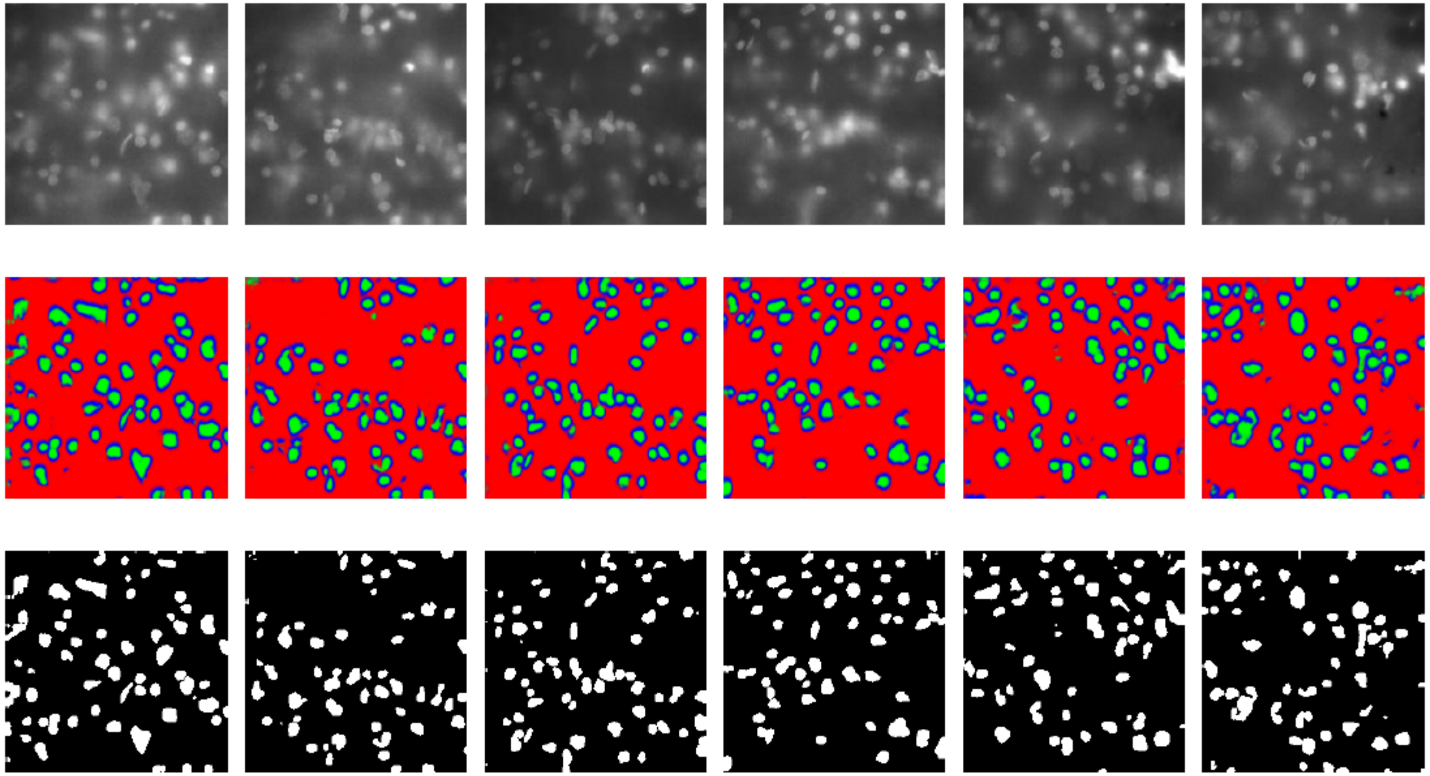
[1] - J. C. Caicedo et al., "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images," BioRxiv, p. 335216, 2019.

Qualitative Results

DAPI Images

Prediction Output
Red: Background
Green: Cell/Nuclei Interior
Blue: Cell/Nuclei Boundary

Binary Map
(Interior + Boundary Classes)



U-Net model generates very reasonable segmentation masks using transfer learning

Proposed Technical Approach
Reduce costly human annotation

Existing model

U-Net Model on nuclei segmentation

→ **Transfer Learning** →

Our datasets

- Large amount of unlabeled nuclei data
- Small amount of labeled nuclei data

→

Results

- Qualitative results
- Quantitative results

Small Labelled Dataset

15 hand-annotated images (LL15)

Annotated with 3 classes:

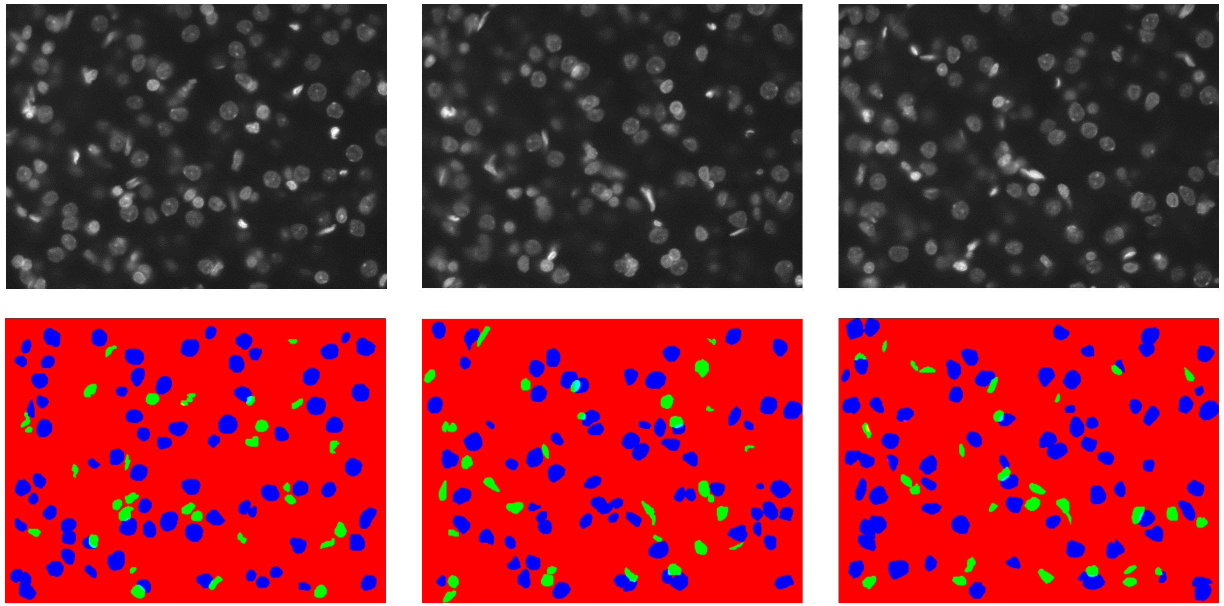
1. Background
2. Glial cell
3. Nuclei

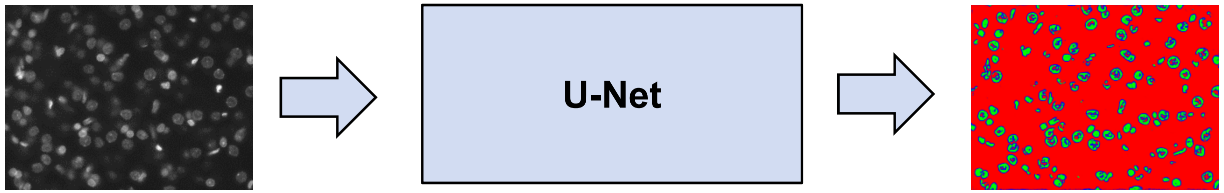
Directly apply U-Net trained on BBB039 dataset on the hand-annotated images

– Combined pixels classified as cell interior and cell boundary to make binary mask

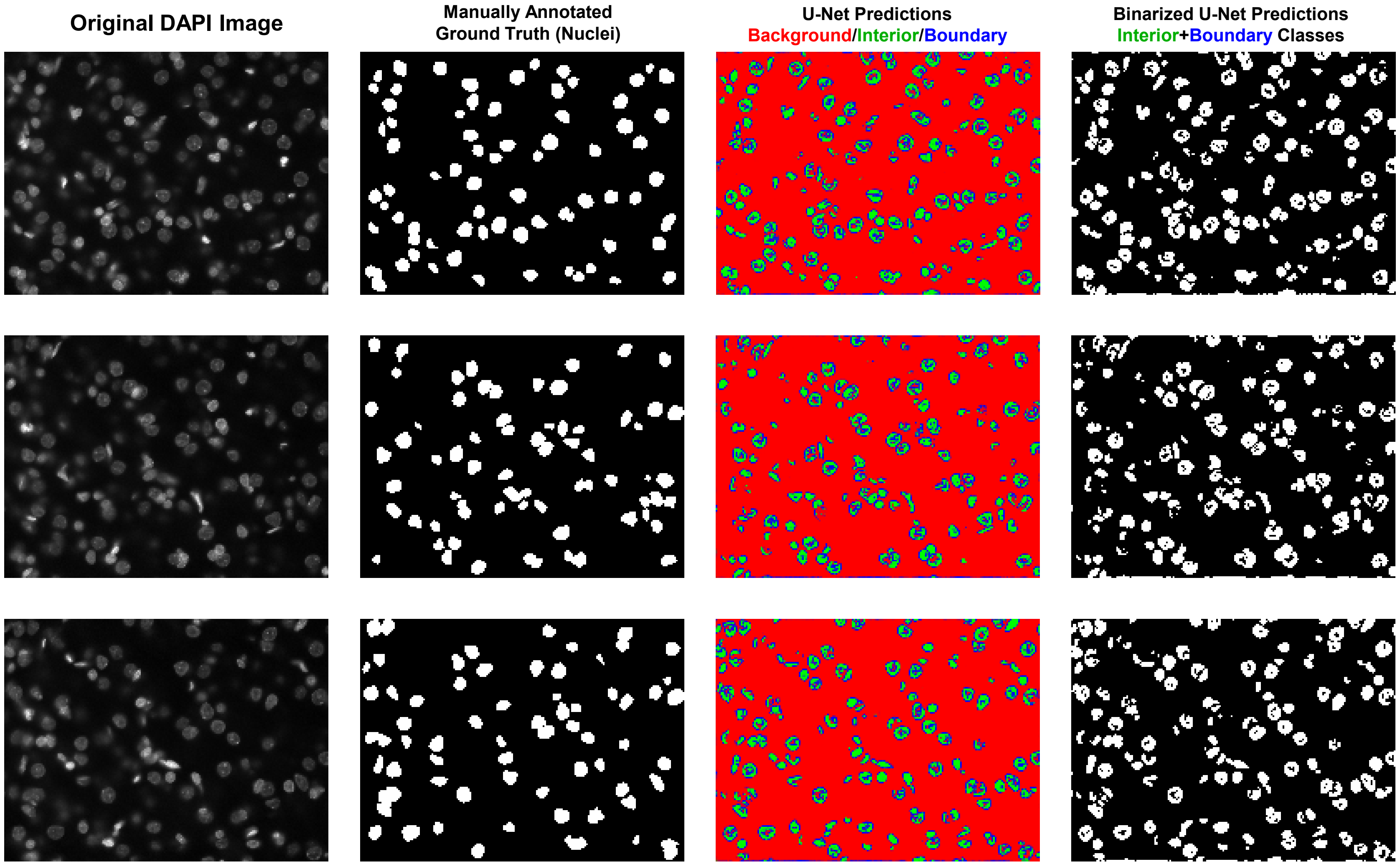
Compute DICE score

– Average over 15 images





Pretraining Results



$$DICE\ Score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

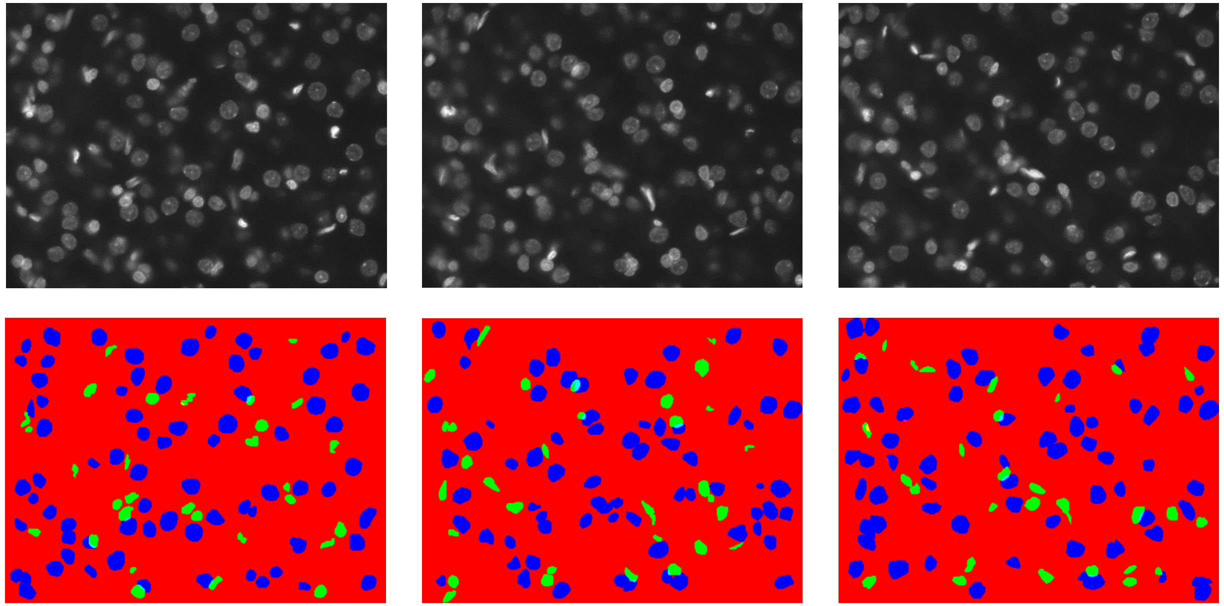
Approach	DICE Score
Automatic (Watershed Segmentation)	0.561
Transfer Learning	0.655

Finetuning Experiment Setup

Can we leverage the labels given in the 15 annotated images to improve performance?

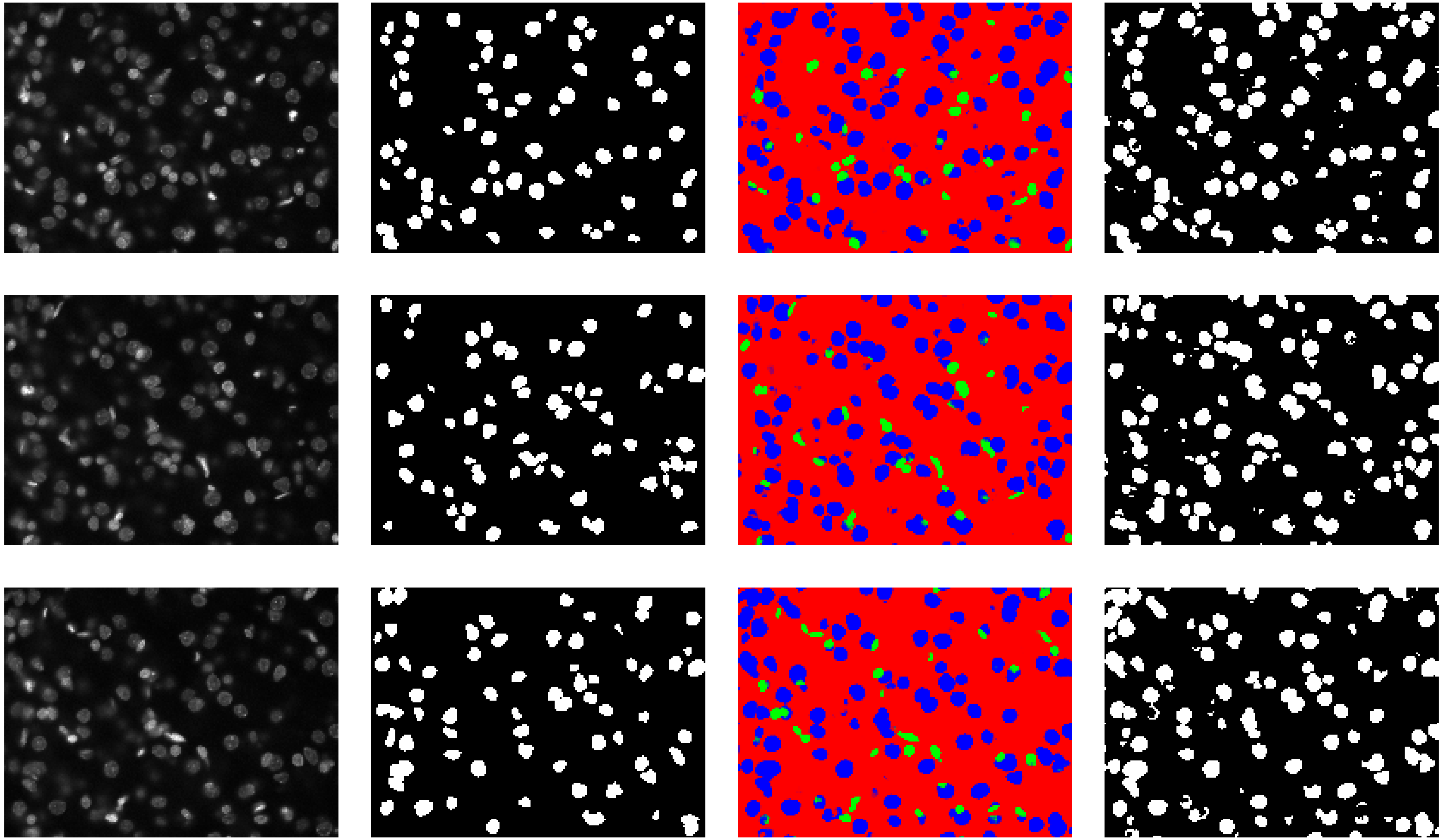
Procedure:

- We perform leave one image out cross validation (train: 14 images, test: 1 image)
- Initialize Model with parameters learned on BBBC039 dataset
- Train for 5 epochs
- Data Augmentation is used
  - Crops, flips, rotations, illumination changes
- Compute DICE score
  - Average across all 15 images



Improved Results

Original DAPI Image | Manually Annotated Ground Truth (Nuclei) | U-Net Predictions Background/Glial/Nuclei | Binarized U-Net Predictions Nuclei Class



$$DICE\ Score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

Approach	DICE Score
Supervised Learning (Baseline)	0.710
Transfer Learning + Supervised Learning	0.733

Comparison of Our Approaches

Approach	Algorithm	Train Dataset	DICE Score
Automatic	Watershed	N/A	0.561
Transfer Learning	U-Net	Broad Institute Dataset (BBBC039)	0.655
Supervised Learning	U-Net	LL15	0.710
Transfer Learning + Supervised Learning	U-Net	Broad Institute Dataset (BBBC039) -> LL15	0.733
Transfer Learning	Mask R-CNN	Kaggle 2018 Science Bowl (BBBC038v1)	0.722

Summary and Future Work

- Automated, high-performance computing approaches are needed to map the brain connectivity
  - Deep learning on brain microscopy data has shown promise
- Human annotation is costly and requires domain expertise
- Develop scalable, learning-based methods
  - Transfer learning on nuclei segmentation shows promising results
  - Future work will make use of conventional image processing results as weakly annotated data
  - Extend to detecting additional classes (e.g., different cell types)