Deep Learning-Based Neuron Detection in Brain CLARITY Imaging

Final Report

Student: Prerna Singh

Class: Computer-Integrated Surgery II (601.455), Spring 2020

Instructor: Dr. Russell Taylor

Mentor: Dr. Jeremias Sulam

Objective: To develop a 3-dimensional convolutional neural network (CNN) that can predict, with increased precision and accuracy when compared to other models, how many fluorescent neurons are present within a section of a brain imaged with CLARITY.

Abstract: We propose a six-layer 3-dimensional convolutional neural network (CNN) to count the number of fluorescent neurons present within a section of a brain imaged with CLARITY, a critical tool for the connectome project[5], enabling the exploration of important neuroscience questions. Our results, cross-validated over different splits show accuracy scores consistently greater than 95% with area under the ROC curves consistently greater than .9. Given the capability to predict with greater than 95% accuracy, we believe that deep convolutional neural networks are capable of detecting neurons in CLARITY images. Potentially, a similar model could also be used for other biomarker detection tasks.

Background

Clear lipid-exchanged acrylamide-hybridized rigid in-situ-compatible tissue hydrogel (CLARITY) imaging enables "lossless high-resolution brain-wide imaging" through the use of tissue transparency techniques.[1] The CLARITY imaging technique is unique in that it makes brain tissue transparent, allowing for a complete 3-dimensional representation of the brain in a 2-dimensional scan. This tissue-hydrogel technique allows for imaging of the most internal parts of the brain, giving a new perspective on the biology of the brain.[2]

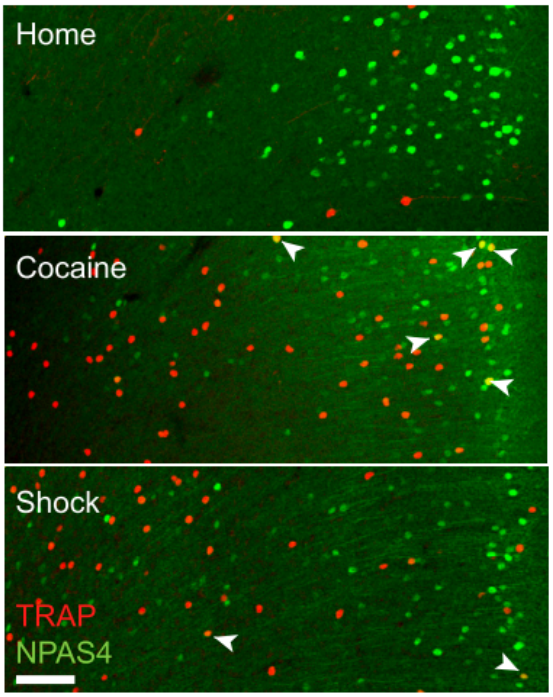

Figure 1: Image from Wiring and Molecular Features of Prefrontal Ensembles Representing Distinct Experiences, Ye et al., 2016, Cell 165[1]

Figure 1 depicts the fluorescence where a specific antibody is binding to a cell in a mouse prefrontal cortex. The different colors represent different cell types: TRAP and NPAS4, which can all be differentiated in these scans.[1]

More generally, specific antibodies can be engineered to bind to cells and cause fluorescence, based on which cells scientists are studying within the brain. Being able to look at the brain in three dimensions allows scientists to focus on certain structures without losing focus of the whole brain in the background by enabling lossless high-resolution brain-wide imaging.

Motivation

CLARITY is a critical tool for the connectome project[5], which looks at how neural cells fire together, within a hierarchy of neural activation patterns in the brain. The overall goal of the connectome project is to develop a comprehensive map of neural connections in the brain. CLARITY can give a lot of insight into that by seeing how different parts of the brain activate and fluoresce with neurons in other locations in the brain, which has not been possible given the intrinsic limitations of previous kinds of imaging, which give only a 2-dimensional picture of cells within the brain.

CLARITY can also help in the understanding of neurotransmitter pathways by using antibodies that bind to neurotransmitters, their precursors and end products, and tracing their location in the brain over time. Lastly, CLARITY can help in studying neurological diseases, by focusing on

diseased or damaged structures in the context of the brain as a whole, where previously, only small 2-dimensional sections of a specific brain structure could be studied like this.

The CLARITY scans used in this project are focused on the connectome project. The experiments try to understand how different cell types in the brain link their activity and pathways. Through CLARITY scans, cellular logic and behavior can be understood through cell-type specific excitation throughout the brain.[1]

In any of these different applications of CLARITY, processing each scan usually entails annotating and counting cells by hand. These images can be fairly large. For instance, in the data set used in this project, there are ~1500 slices per brain and each slice is ~5.5 million pixels. In practice, this can be very time consuming and inefficient, given the amount of information that can be extracted by these scans. Previous automatic methods used to solve this problem include filtering, template matching, and blob detection, which had a maximal accuracy of ~59%. These are all non-learning methods, which begs the question whether a deep learning approach would perform better. Because a cell spans across multiple slices as a three-dimensional (3D) object, development of a 3D convolutional neural network (CNN) to count the fluorescing cells will streamline the processing of CLARITY scans, allowing for critical information from these scans to be available almost immediately.

I investigate the development and use of 3-dimensional convolutional neural networks to detect fluorescent neurons present within a section of a brain imaged with CLARITY and compare its performance with other methods.

Technical Approach

The project workflow is split into the following main components:

1. Design and Implementation of 3-dimensional CNN Model
2. Training and validation of 3-dimensional CNN training on reduced dataset (1 brain)
3. Large scale training on complete dataset (8 brains) and validation

Design and Implementation of 3-dimensional CNN Model

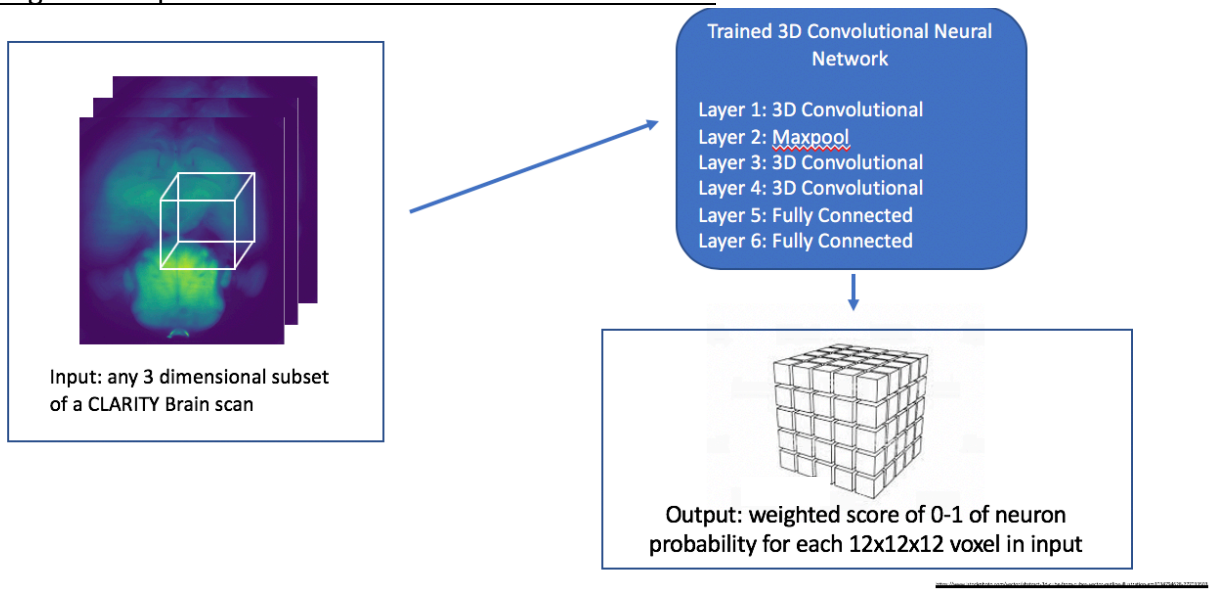

Figure 2: The inputs and outputs of trained 3D CNN.

The 3D CNN Layers proposed are based on the paper "*Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks*" that detects cerebral microbleeds in 3D brain scans and outperformed previous attempts at this problem with increased accuracy and sensitivity and fewer false positives.[3] Figure 2 shows the structure of the 3D CNN. The CNN has three convolutional layers, one maxpool layer, and ends with two fully connected layers. The advantage of using a 3D CNN is that it can take advantage of the spatial contextual information to extract more representative high-level features for detecting these neurons.[3]

The proposed 3D CNN can take a 3D input of any size greater than 12x12x12 pixels and reports a score ranging from 0 to 1 whether it is a fluorescing neuron (1) or not (0). From here, a sensitivity threshold will be determined based on optimization of performance metrics (precision, accuracy, sensitivity, recall, etc.) that will lead to a binary decision of fluorescing neuron (1) or not (0).

Training and validation of 3-dimensional CNN training on reduced dataset

The complete dataset for this project are CLARITY images for eight full brains. It contains ~1500 two-dimensional images, with 2560x2160 pixels each, taken at different depths of the brain. These eight brains contain annotations for the coordinates of *some* of the fluorescing cells, totaling approximately 2000 annotated cells. *Note that not all the fluorescing cells are annotated.*

Figure 3 portrays the workflow of training the CNN. Based on my empirical measurements and previous work, it was determined that a voxel size of 12x12x12 pixels will contain the entire cell. Thus, 12x12x12 voxels are used for training, with both positive and negative controls. Figure 4 shows an example of a positive cell region and a non-cell region. Voxels that contain a fluorescing cell were used as positive training data for the CNN. To generate negative samples (non-

fluorescing cell) to train the CNN, ten times the amount of non-cell regions were generated by randomly sampling 12x12x12 regions that are within 6-20 pixels of regions centered on the cell annotations. This was done because the number of fluorescing cells within the brain are actually a very small fraction of all cells.

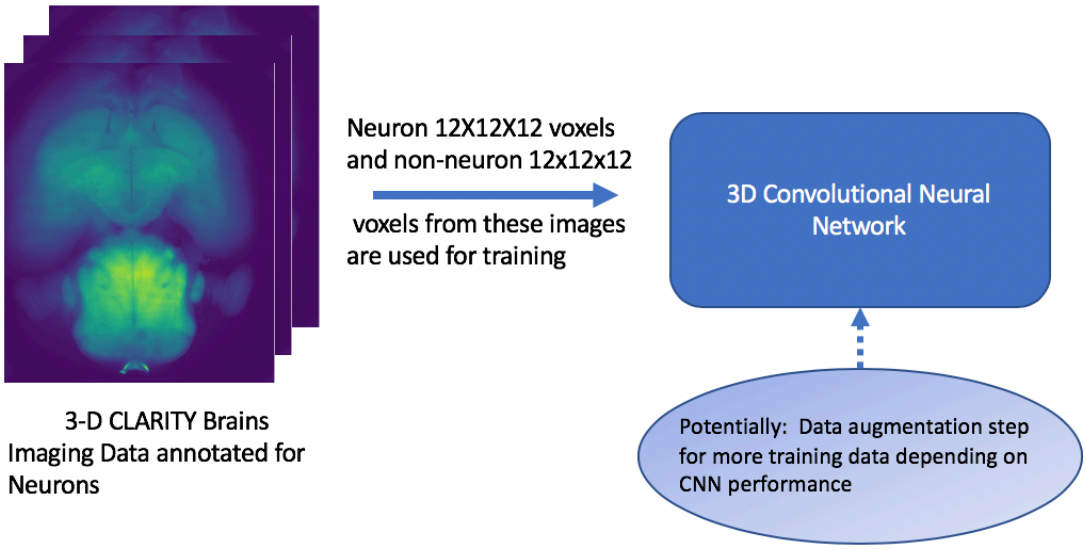

Figure 3: Workflow of training the 3D CNN. 12x12x12 voxels are used for training, with both positive and negative controls.

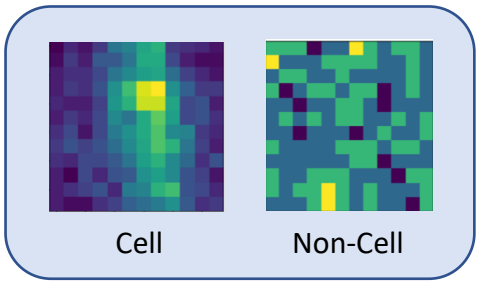

Figure 4: The image on the left shows an example of what a slice of a positive cell region should look like. The image to the right shows an example of a slice of a negative training sample.

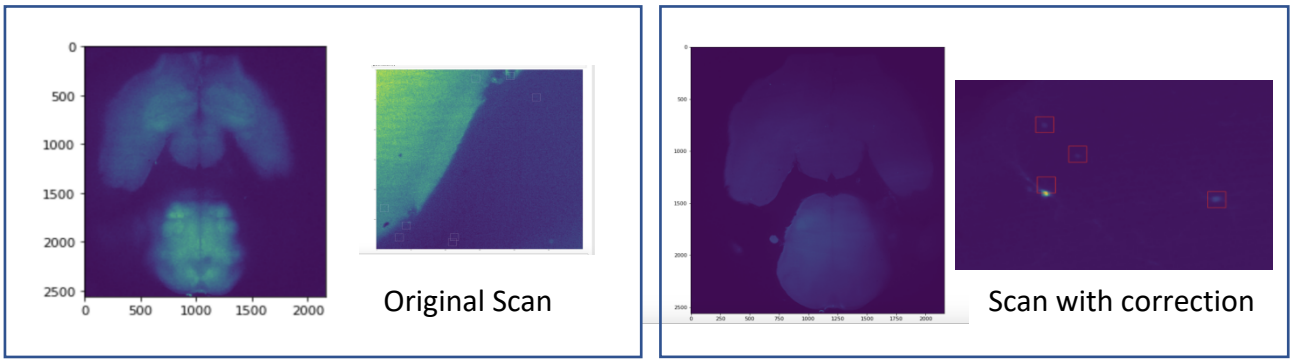

Figure 5: The example on the left shows a CLARITY scan before the illumination correction. The example on the right shows the results of a brain illumination correction. We can see that it is much easier to identify the fluorescent cells in the scan post illumination correction

Due to the variation in brain intensity in the scans, before extracting any data for training, a brain illumination correction filter, N4ITK[6] was applied to the data. Without the N4ITK filter, it was quite hard to visualize the fluorescent cells, even with the human eye, since there was so much variance in the background. By applying a brain illumination correction filter to the data, we can see that it becomes much easier to distinguish the fluorescent cells. Figure 5 shows an example of using the illumination correction filter.

Large scale training on complete dataset (8 brains) and validation

We used all eight brains for training the final model. However, in order for the validation to be unbiased, two different validations were performed, by excluding some training data and using it for testing. The first is coarse validation, where metrics such as area under the ROC (AUC) and PR (APR) curves, precision, recall, and accuracy will be reported over random splits of training vs. testing data. Each of the splits were also retrained with some false positives and all of the false negative results. The second validation is full cross validation, where the same metrics will be reported and averaged over four folds. In each fold, six brains were used for training and the remaining two for testing.

Results

Table 1 shows the results of two splits of coarse validation with retraining on some false positives and false negatives and achieves F1 score ranging from 0.7251 to 0.7784 with accuracies exceeding 0.96 and precision ranging from 0.95 to 0.97. Figure 6 shows the ROC and PR curves for each of the four splits.

Table 2 shows the cross-validation results averaged over four-fold testing yielding F1 score of 0.737 with accuracy of 0.9614, precision of 0.9649 and recall of 0.5963. The AUC is 0.916 and PR is 0791.These are excellent results. Figure 8 shows the ROC and PR curves for each of the four folds.

Figure 7 shows the false positives from Split 2 retrained. In this case, 14 of 19 regions (the bottom two rows) are cell regions. Because of this, the adjusted number of false positives for Split 2 retrained is 5.

Table 1: The table shows the results of two splits of coarse validation with retraining on some false positives and false negatives. The Adjusted FP column gives the number of actual False Positive, since some of the False Positives were actually cell regions that were classified correctly.

Coarse Validation Results

Split	F1 score	Accuracy	Precision	Recall	FP	FN	AUC ROC	AUC PR	Adjusted FP
1	0.7612	0.9647	0.9867	0.6203	8	77	0.924	0.817	4
1 retrained	0.7784	0.9664	0.9733	0.6485	17	49	0.935	0.840	9
2	0.7730	0.9657	0.9701	0.6425	12	61	0.920	0.805	8
2 retrained	0.7251	0.9596	0.9493	0.5865	19	76	0.909	0.775	5

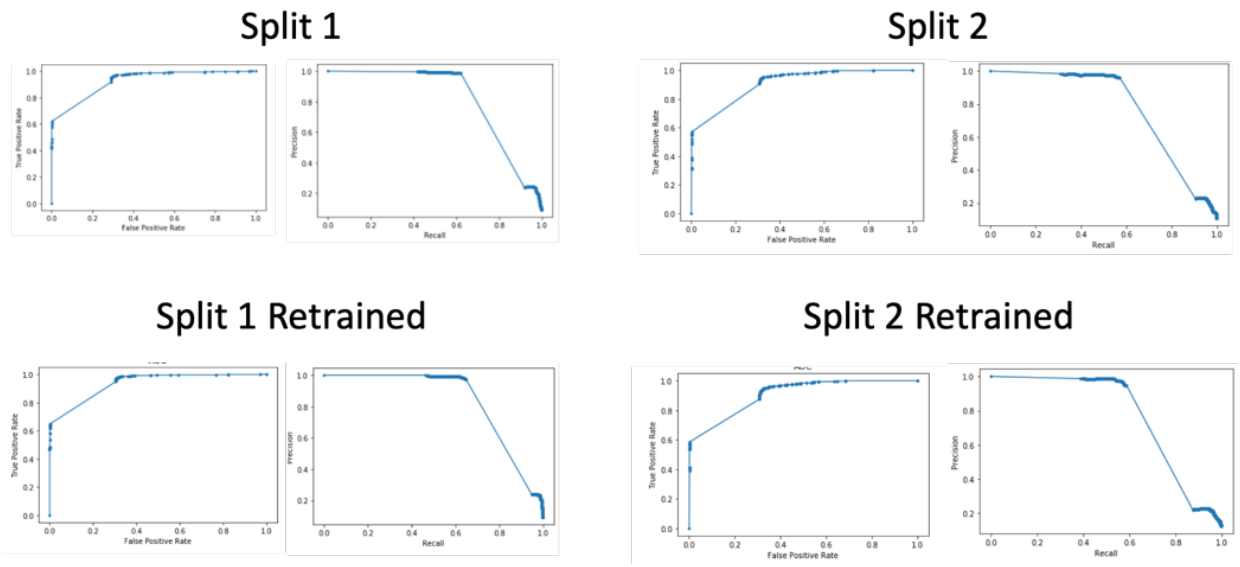

Figure 6: For each of the training cases, the ROC and PR curves are shown. The area under all of the curves are shown in Table 1.

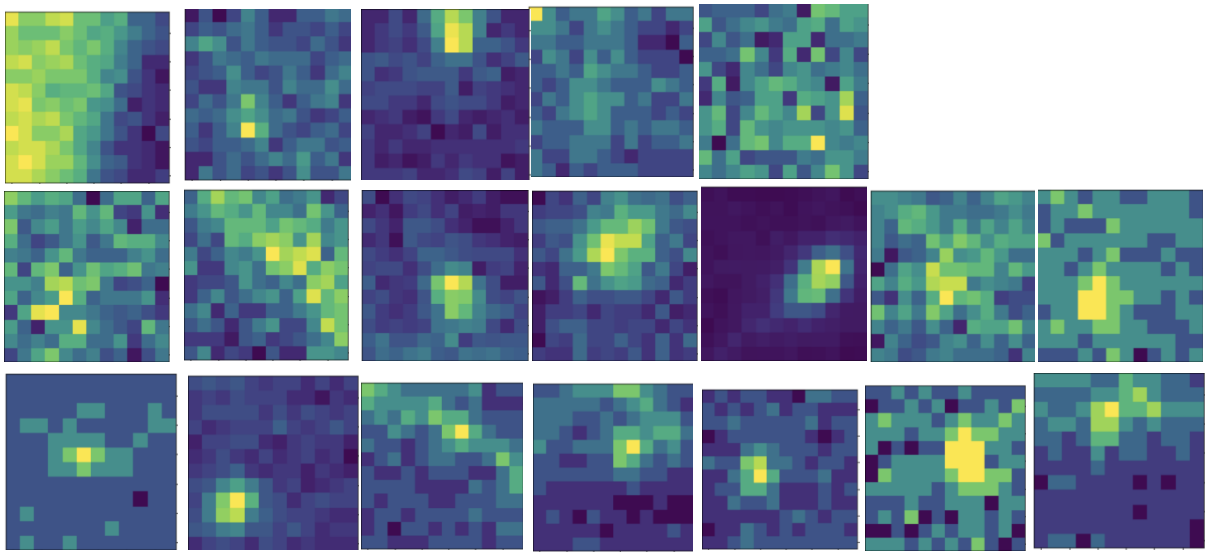

Figure 7: This shows the false positives from Split 2 retrained. In this case, 14 of 19 regions (the bottom two rows) are cell regions. Because of this, the adjusted number of false positives for Split 2 retrained is 5.

Table 2: The table shows the results of cross-validation. The Adjusted FP column gives the number of actual False Positive, since some of the False Positives were actually cell regions that were classified correctly.

Cross Validation Results

Cross-Validation	F1 score	Accuracy	Precision	Recall	FP	FN	AUC ROC	AUC PR	Adjusted FP
Averaged	0.7370	0.9614	0.9649	0.5963	14.25	62	0.916	0.791	3.5

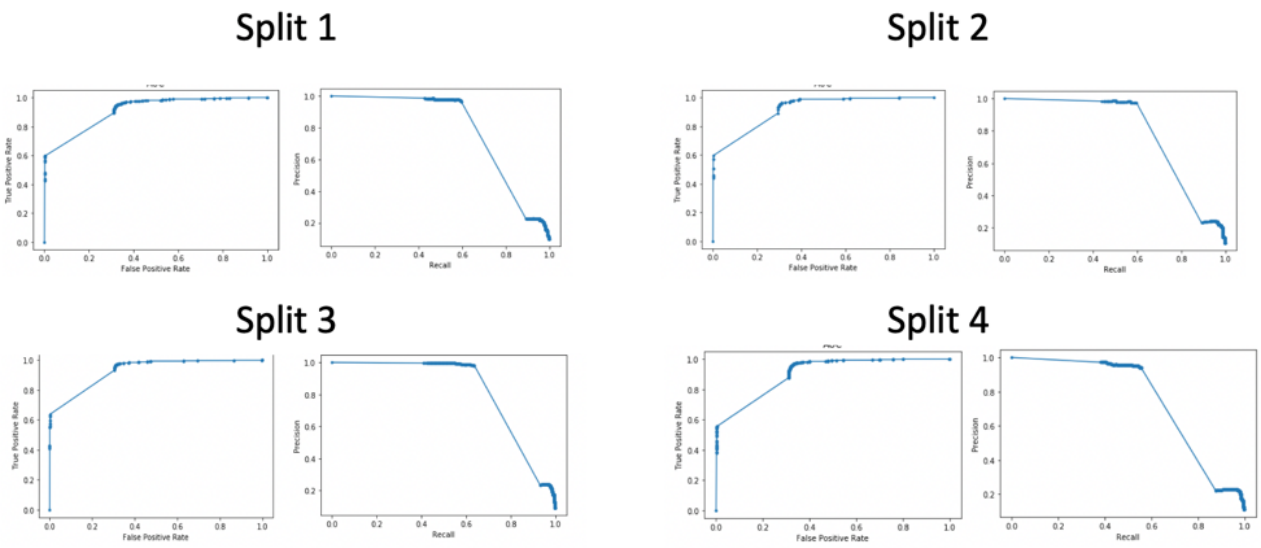

Figure 8: For each of the training cases for the cross-validation, the ROC and PR curves are shown. The averaged area under all of the curves are shown in Table 2.

Discussion of Results and Significance

Given the results we can conclude that the proposed model actually performs even better than the statistics shown. In Figure 7, the false positives from Split 2 retrained are displayed. 14 of these 19 voxels do actually seem to contain a cell. This can be attributed to the fact that the training data had only a fraction of annotated cell regions and there were many more cell regions that were not annotated. In the way the non-cell training data was generated, some of the non-cell regions actually capture cells, and the model does a good job of classifying those as such, based on the adjusted FP column in Table 1, which accounts for the incorrectly labeled training data. Given this, we can assume that all of the metrics used to quantify the quality of the 3D CNN are actually better than what is shown in the tables, ROC, and PR curves.

The proposed 3D CNN model surpasses the accuracy of the previous model significantly. The results of the coarse and cross-validation show accuracy scores consistently greater than 95% compared to 59% by previous methods[7]. This can lead to major breakthroughs in neuroscience research. To be able to analyze any brain scan, identifying neurons is critical. Given the capability to predict with greater than 95% accuracy, we believe that deep convolutional neural networks are capable of detecting neurons in this kind of imaging technique. Potentially, a similar model could also be used for other biomarker detection tasks.

Computational Infrastructure

Due to the immense computational requirements necessary to train a 3-dimensional convolutional neural network, MARCC (Maryland Advanced Research Computing Center https://www.marcc.jhu.edu/) resources were utilized. Blue Crab is the main cluster at MARCC

with over 23,000 cores (June 2018) and a combined theoretical performance of over 1.4 PFLOPs. 72 GPU nodes were used to complete the training and validation of the proposed model. The GPU nodes are Dell PowerEdge R730 servers with dual Intel Haswell Xeon E5-2680v3 (12 core, 2.5 GHz, 120W), 128 GB of 2133 MHz DDR4 RAM. (AVX frequency: 2.1GHz) and two Nvidia K80s per node. Each training run took 4-6 hours with 72 GPUs.

Deliverables

The minimum and expected target deliverables were achieved in this project. Moreover, majority of the maximum target deliverables were also achieved as the current project results are significantly better than previously published results. I am in the process of further validation of the results with neuroscientists before the academic paper can be completed. Below are the specific deliverables listed and their status.

Due to the usage of MARCC for COVID-19 related activities, it was hard to get the required computing resources for the project in a timely manner. This was ultimately the biggest bottleneck in the progress of the project.

Minimum: **[COMPLETED]**

- Trained 3D CNN model (pytorch) for neuron detection
- Report in Jupiter notebook – explains training of CNN and predictions

Expected: **[COMPLETED]**

- Robust and validated trained model (pytorch) for neuron detection

Maximum: **[PARTIALLY COMPLETED]**

- Academic paper describing packaged model, training, validation
- Surpass baseline accuracy of previous models by a significant margin **[COMPLETED]**

Management Plan

The management plan included meetings with my mentor, Dr. Sulam and software development process. These are itemized below.

- Meetings with Dr. Sulam as necessary
- Code stored on github and MARCC
- Code backed up to github every week and more often with substantial improvements
- Communication through Slack/email

Code

The entire software source code is stored in a linked github repository on Wiki (https://ciis.lcsr.jhu.edu/dokuwiki/doku.php?id=courses:456:2020:projects:456-2020-06:project-06).

Bibliography and Reading List

1. L. Ye, W. E. Allen, K. R. Thompson, Q. Tian, B. Hsueh, C. Ramakrishnan, A.-C. Wang, J. H. Jennings, A. Adhikari, C. H. Halpern, I. B. Witten, A. L. Barth, L. Luo, J. A. Mcnab, and K. Deisseroth, "Wiring and Molecular Features of Prefrontal Ensembles Representing Distinct Experiences," *Cell*, vol. 165, no. 7, pp. 1776–1788, 2016.
2. Karl Deisseroth, "Building a See-through Brain- A New Experimental Approach at the Interface of Chemistry and Biology," *Scientific American,* 2016
3. Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.
4. C. Magliaro, A. L. Callara, N. Vanello, and A. Ahluwalia, "A Manual Segmentation Tool for Three-Dimensional Neuron Datasets," *Frontiers in Neuroinformatics*, vol. 11, 2017.
5. "NIH Launches the Human Connectome Project to Unravel the Brains Connections," *PsycEXTRA Dataset*, 2009.
6. N. J. Tustison, B. B. Avantis, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 Bias Correction", *IEEE Transactions on Medical Imaging,* June 2010, 29(6), pp. 1310-1320.
7. Vikram Chandrasekhar, This method was developed in Dr. Sulam's lab and based on template-matching.