12-2018

Differentiation of Neurons and Glia for Use in Cellular Connectomics

Jacob T. Brettin
University of Tennessee, Knoxville, jbrettin@vols.utk.edu

Differentiation of Neurons and Glia for Use in Cellular Connectomics

Jacob Brettin

Neuroscience Program, Univ. of Tennessee

Introduction

As the field of neuroscience advances, it is becoming increasingly clear that the structure of the brain is directly associated with its function. Moreover, by investigating how alterations in the brain's structure affects its functions, researchers are finding novel ways of detecting and treating various neurological and psychiatric disorders (Beyer et al., 2013). By mapping the brain's neural pathways, researchers can reconstruct digital images of the brain for a variety of purposes. Due to the significance of these maps, the NIH founded the Human Connectome Project with the purpose of characterizing neural connectivity, function, and variability through the use of digital reconstructions of the human brain, also known as connectomes ("Connectome", 2018). This paper seeks to provide a background on the field of connectomics as well as to detail our current work in the construction of connectomes capable of neuron-glia differentiation.

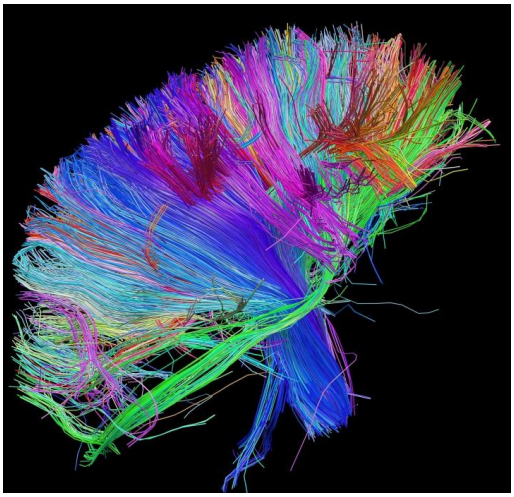

Figure 1. A connectome of white matter fiber architecture produced by the Mark and Mary Stevens Neuroimaging and Informatics Institute at USC. Digital maps such as these help scientists uncover the relationship between structure and function in the brain. This connectome and others can be viewed at http://www.humanconnectomeproject.org.

Macroscale Connectomics

Currently, there are two primary types of connectomes—macroscale and cellular. Macroscale connectomics has the advantage of being non-invasive and is finding increasing use within the medical field. Applications of macroscale connectomics include visualizing anatomical structures and abnormalities, observing changes in the brain, and for image-guided surgical planning (Despotović, Goossens, & Philips, 2015). Currently, macroscale connectomes are constructed using magnetic resonance imaging (MRI). MRIs function by aligning resonating hydrogen atoms using a magnetic field and measuring the radio waves emitted by the atoms as they return to their resting state. The intensity of the signal is dependent on the concentration of hydrogen atoms which in turn is dependent on the type of tissue present. This intensity is measured on a grayscale to produce a cross-sectional image (Berger, 2002). When multiple cross-sectional images are produced in sequence, a 3D image of the brain is produced.

Once a 3D volume is obtained, it is subjected to classification and segmentation. In brain MRIs, image elements are typically classified as either white matter, gray matter, or cerebrospinal fluid. Additional subclassification can be applied depending on the purpose of the connectome. Once image elements are classified, they are segmented into regions of similarly attributed elements to provide a semantically meaningful connectome (Despotović et al., 2015). In the medical field, MRIs can be segmented to display a variety of structural abnormalities, from tumorigenic regions to various types of head trauma-induced hemorrhages (Gong et al., 2007). In addition, segmentation of MRIs is used to depict axonal projections for research-based purposes (Mori et al., 1999).

Cellular Connectomics

Cellular connectomics provides an alternative view of neural connectivity and is the primary focus of our research. Due to the high resolution necessary to accurately map individual cells and their synapses, non-invasive techniques have less practical use in cellular connectomics (Kasthuri et al., 2015). As a result, these connectomes are not used clinically, but rather in a laboratory setting. Currently, image datasets used in cellular connectomics are primarily produced via serial electron microscopy (EM). In serial EM, tissue is sliced into slices 20-30 nanometers in thickness, each of which is then imaged in series. The digital image series is then compiled into a 3-D representation of the unsliced brain sample (Kasthuri et al., 2015). The images produced by serial EM have extremely high resolution, and the slices used in the imaging typically range from 20-30nm in thickness (Dyer et al., 2017). Subsequent segmentation of these volumes can be done manually or via algorithm to provide a variety of data, from neuron densities to tracings of entire neural pathways. Classification types for image elements are abundant and depend on the aspects of neuronal connectivity being studied; some connectomes may focus on a particular type of neuron, while others, such as the mouse neocortex volume described in Kasthuri et al., classify multiple cellular and sub-cellular components (2015).

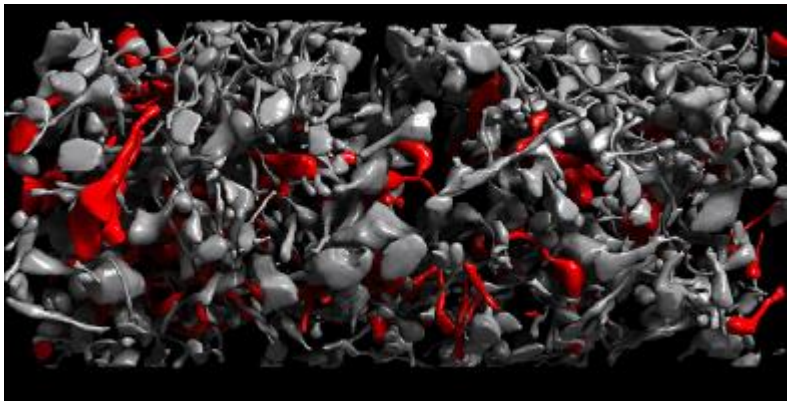

Figure 2. Figure excerpt from **Kasthuri et al**. Digital reconstructions of the brain are allowing researchers to examine details regarding the connectivity of cells that were previously unobtainable.

Unfortunately, EM segmentation is very costly in terms of time and storage—a cubic millimeter of brain tissue at 20nm resolution requires approximately two petabytes of storage and takes about three months to process (Dyer et al., 2017). Recently, this time and storage problem has been solved by using X-ray tomography in place of EM. Synchrotron X-ray microtomography is capable of producing cross-sectional images of centimeter-scale brain samples, thereby eliminating the need for imaging individual thin slices as seen in EM techniques. As such, the speed at which large-scale image datasets are produced is greatly increased (Dyer et al., 2017). For this reason, X-ray tomography is seen as the more viable option for large scale brain mapping.

Use of neural networks in connectomics

Classification and segmentation of image datasets can be done manually by a trained observer. While this may be the most accurate way of tracing neural pathways, manual tracing of connectomes quickly loses feasibility as the size of the datasets increases. As mentioned above, a reasonably sized connectome, only one cubic meter in volume, contains two petabytes worth of information (Kasthuri et al., 2015). In order to make sense out of this amount of data, researchers have turned towards image processing algorithms. These algorithms greatly reduce the time it takes to classify and segment large image datasets. While they are generally not as accurate as manual tracing, this flaw is frequently compensated by manual checking of the algorithm, a task much less daunting than physically classifying and segmenting the entire dataset by hand (Egmont-Petersen, De Riddler, & Handels, 2002).

Image processing algorithms are able to address a variety of problems encountered in the analysis of large image datasets. Egmont-Petersen et al. outline the multiple processes that

common neural networks use to produce meaningful output from an image (2002). These tasks include 1) preprocessing of the image to smooth edges and reduce noise, 2) data reduction and extraction of features of interest, 3) segmentation of similarly classified regions, 4) object detection and classification, and 5) image understanding, which outputs semantically meaningful information (Egmont-Peterson et al., 2002). Neural networks can be designed that address one or more of these functions. In the field of connectomics, segmentation and object detection/classification neural networks are of extreme importance and are responsible for the construction of 3-D brain maps.

In order to automatically construct a connectome, a machine algorithm must be capable of completing the following tasks:

1) Preprocessing: Preprocessing algorithms are a means of reducing the number of extraneous variables within the dataset. In the case of macroscale connectomics, the skull is almost invariably removed as it does not typically contribute to the connectome (Gong et al., 2007). In the study of hemorrhages by Gong et al., gray matter is also removed (2007).

2) Feature extraction: Any image can be broken down into multiple feature elements such as grayscale value, line segment orientation, and area (Gong et al., 2007). Objects within an image have specific values for these feature elements; based on the features present in an image region, the probability of the object's presence within that region can be determined (Indolia et al., 2018).

3) Classification: Once an object is detected by the algorithm, it is often necessary to know what this object is. Based on an object's associated features, the algorithm can predict the likelihood that the object is associated with a specific classifier. Examples of

classifiers used in connectomics include cell bodies, axonal projections, dendritic spines, as well as specific cell types (excitatory/inhibitory; neuronal/glial) (Kasthuri et al., 2015).

4) Segmentation: A connectome is comprised of not a single image, but a group of images representing 3-D space. As such, one cell will be present in multiple images that are inputted into the algorithm. The algorithm therefore must be able to determine whether an object within an image is associated with an object in adjacent images. By segmenting image objects together, a 3-D representation of the single object is created (Kasthuri et al., 2015).

Convolutional Neural Networks (CNNs) are commonly used to construct connectomes from image datasets. These networks are biologically inspired and take advantage of the hierarchical system used in the mammalian visual pathway wherein complex features are detected from the presence of many simpler features (Fieres, Schemmel, & Meier, 2006). The first hidden layer of a CNN divides the image into small regions and extracts important simple features regarding these regions such as line segment orientations. Subsequent layers then piece these simple features together to detect increasingly complex shapes and eventually recognize whole objects within the image. Fieres et al. outline the steps a CNN takes towards whole object recognition as well as the biological foundations underlying this type of neural network (2006). This particular type of neural network is useful for constructing connectomes because it requires less computational effort than traditional neural networks through the use of their hierarchical system (Fieres et al., 2006).

Objective of Research

Our research is ongoing and focuses on the construction and implementation of cellular connectomes. The study that is the focus of this paper has three main objectives:

1) Identify visible differences in EM images of neurons/glia: The importance of glia in a variety of neuronal functions have only recently been recognized **(**Almad & Maragakis, 2012). While images of glial cells can be identified among neuronal cells by a trained eye using qualitative identifiers such as shape, there is a surprising lack of known quantitative variables that can be used to distinguish between cell types in image datasets. Quantitative variables are necessary for automatic differentiation of cell types using neural networks, as computer algorithms rely on numbers to create predictive models. The primary goal of our study is to test the validity of the ratio of cell nuclei-to-body diameter as a quantitative identifier for neural and glial cells in a 3-D volume produced from serial EM.

2) Translate EM study to X-ray images: Because of the benefits X-ray tomography provides in terms of speed and size, a secondary goal of our work is to determine the translational ability of our work using EM datasets to other datasets produced from X-ray imaging techniques. This will promote the transition of future studies from EM to X-ray data, which we anticipate will become the primary method of producing images used in cellular connectomics.

3) Begin creating algorithm for automatic segmentation: Pending the success of our initial studies, we aim to begin designing an image-processing neural network capable of classifying and segmenting neuronal and glial cells to construct a connectome that distinguishes between cell types. By applying the results of our study to the construction

of connectomes, we hope to learn more about how the organization of glial cells affects brain functions.

Procedure

Analysis of EM Data

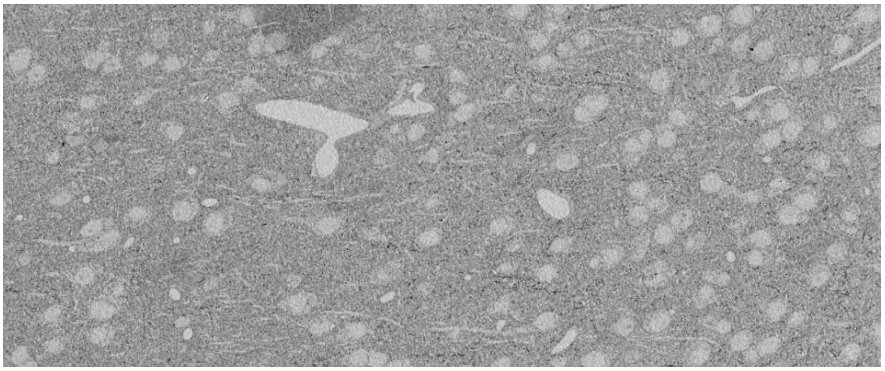

Figure 3. An EM image of a slice of S1 mouse tissue being viewed through the VAST program (Kasthuri et al., 2015).

In our search for quantitative variables capable of distinguishing neural and glial cells, we chose to analyze the maximal diameter of each cell's body in relation to the maximal diameter of the cell's nucleus[1]. A portion of mouse cortical tissue from the S1 region was partitioned into slices using the automatic tape-collecting ultramicrotome (ATUM) and imaged with a scanning electron microscope as detailed by Kasthuri et al (2015). The resulting images were compiled into a single dataset. In order to perform the necessary manual annotations and measurements, the Volume Annotation and Segmentation Tool (VAST) program was utilized (Kasthuri et al., 2015). This user interface system allows the user to scroll through each image, or slice, of the 3-D volume and make drawings or annotations on either the individual slice or on multiple slices. For each cell, a line was drawn across the diameter of the cell on the slice where the cell appeared to have the greatest diameter. This process was repeated for each cells' nucleus. To minimize bias, cells of both types were annotated without prior knowledge of which cells were glial or neuronal. Once the annotations were made for both cell bodies and nuclei, each cell was subsequently classified as either neuronal or glial.

[1] The research included in this thesis was conducted under the supervision of Narayanan Kasthuri, MD, PhD at the Argonne National Laboratory.

Another feature of the VAST program is the inclusion of an (x,y,z) coordinate system that assigns each pixel in *z* slice an *x* and *y* value. This feature was utilized in the calculation of the length of the diameters. Using the (*x*,*y*) coordinates of the pixels at the terminal ends of each line in the distance formula $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$, their diameters were calculated in pixels. Using the calculated distances, histograms were made of the distribution of cell body diameters, nuclei diameters, and the ratio of the two were made to assess whether any trends could be found.

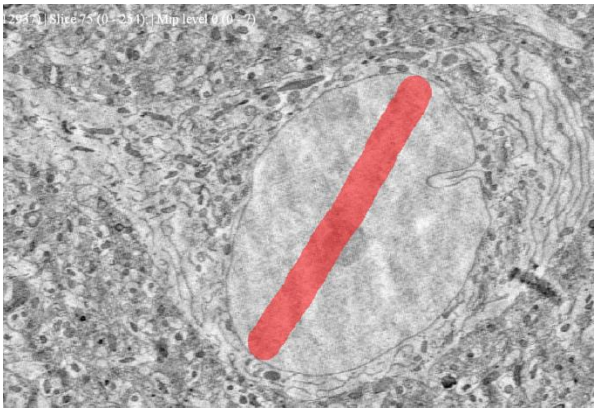

Figure 4. Preliminary annotation of a cell using VAST. Pixels at each end of the annotation were used to calculate diameters of cells and their nuclei.

X-ray Data Comparison

As advances in 3-D image construction and analysis are made, researchers have been pursuing the construction of whole-brain scale cellular connectomes. In order to compensate for the tremendous amount of data and processing this will require, the transition from utilizing EM imaging to X-ray imaging in the construction of cellular connectomes is being pursued (Dyer et al.). As such, before a computational model is constructed to determine cell type, we wanted to

look into whether the values calculated in our EM data will be similar to a different brain sample of the S1 region mapped via X-ray. To do this, a neural network capable of measuring X-ray cell body diameters was used on a large sample of S1 tissue[2]. Meanwhile, the EM values were converted from pixels to micrometers so that the X-ray and EM values would use the same units. While the algorithm used is capable of determining cell body diameters, it cannot give values for the nuclei diameters. To determine whether the nuclei are of comparable diameters and to confirm the accuracy of the algorithm, manual annotation will be required.

Algorithm design

Although more data will be needed before we can begin writing and training an algorithm for automatic segmentation and classification of neurons and glia, we began planning a general design for our neural network. Fortunately, the Human Connectome Project encourages open-sourcing information regarding connectomes, datasets, and algorithms. As such, we reviewed the literature to determine whether neural networks already available can aide us in our goal of constructing a cell type-based connectome. The details regarding our design and literature findings are detailed below.

Results

Analysis of EM Data

Using a 2 tailed T-test, we can conclude that there is a significant difference between the neuronal ratio and glial ratio ($p = 1.003E-8$). Although some overlap occurs between the ratios of the two cell types, there is much less overlap observed when looking at either the nuclei or cell bodies' diameters. This suggests that looking at either statistic alone would be sufficient to determine with a high degree of accuracy whether a cell is neuronal or glial. In cases where a cell

[2] Special thanks to Eva Dyer, PhD for running the X-ray sample analysis in her lab at Northwestern University.

or its nucleus's diameter falls in the area of overlap, using both distances may be enough to determine the cell type. More samples of the cell could further reduce this overlap to provide a more accurate prediction of cell type.

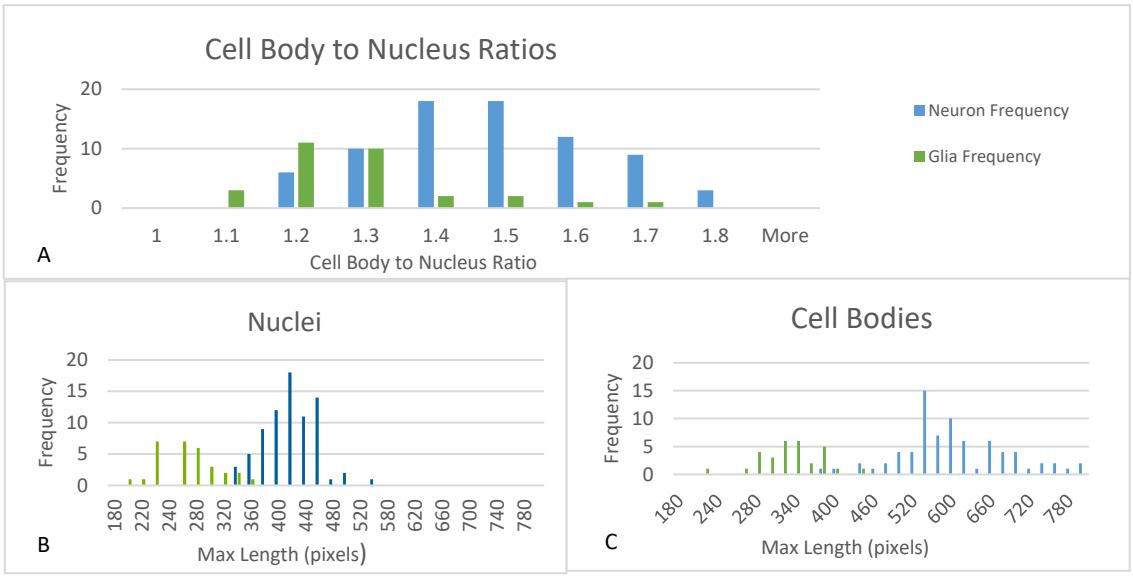

Figure 5. A. Histogram representing the frequency in which ranges of ratios of cell body diameters to nuclei diameters for both neurons and glia were observed. Subset are histograms depicting frequencies of **B.** nuclei and **C.** cell body diameters were observed for neurons and glia.

X-ray Data Comparison

The computer algorithm used to measure and analyze cell diameters found two peaks in its distribution, as seen in Figure 6. These peaks in distribution are of similar size to the peak distributions of cells identified as neurons and glia in EM data. Although manual annotation is still being completed to confirm the algorithm's data, the similarity of the preliminary values suggests that cells in X-ray data can be assigned as either neurons or glia based on the assignment of cells of similar length in EM data. If the manual annotation can also confirm the similarity of X-ray and EM nuclei diameters, then work can be done to begin designing an algorithm capable of distinguishing cell types within X-ray tomography brain samples.

	Peak in Distribution of Glia	**Peak in Distribution of Neurons**
X-ray	8.45 μm	17 μm
Electron Microscopy	9.6 μm	16.8 μm

Figure 6. Peak distribution of cell diameters in EM and X-ray data.

Image Processing Network

The ultimate goal of this project is to utilize neural networks to create a connectome that distinguishes neuronal and glial cells based on the relative size of their nuclei. In order to achieve this goal, the algorithm should be able to complete the following tasks:

1. Preprocessing: To reduce the amount of extraneous data available in future steps, the algorithm should first eliminate blood vessels and extracellular matrix (ECM). This can be accomplished by grayscale detection of blood vessels and texture/gradient-based detection of ECM. Furthermore, grayscale and texture/gradient-based detection of membrane structures is necessary for automatic annotation of cell body and nuclear membranes.

2. Extraction of Features: Once membranes have been identified, several features regarding the membranes should be extracted. Features such as spatial associations with other membranes in the (x,y,z) plane, diameter, grayscale value, and pixel gradient.

3. Classification of membrane type: based on extracted membrane features, membranes in each slice should be classified as cellular or nuclear.

4. Segmentation: By combining associated membranes from each slice of the dataset, a connectome of the cells and their associated nuclei can be produced.

5. Classification of cell type: using the extracted data regarding the major axis length of both types of membranes, a prediction regarding the cell type (neuron vs glia) is made.

Due to the overlap in the distributions of cell diameters, nuclei diameters, and the ratio between the two for neurons and glia, a neural network capable of automatically distinguishing between neurons and glia will need to consider all three distributions when determining the cell type. However, we acknowledge that even when using all three measurements in conjunction with each other, there will likely be situations in which a cell's type cannot be determined with a high degree of probability. As such, it will be useful to include a confidence output, so that cells automatically classified with a low degree of confidence may be flagged to be checked manually. In addition, the inclusion of a confusion matrix will aid in the identification of falsely classified cells. Once automatic segmentation is achieved, we can modify the neural network to include other identifying information such as spatial distributions to further increase the confidence of correct segmentation.

In addition to outlining the major tasks we believe are necessary for our algorithm to accomplish, we also searched the available literature for currently existing neural networks that may be capable of learning to create cell type-based connectomes. RhoANA is an open-sourced semi-automated segmentation tool that has been shown to adequately classify membrane structures in large EM datasets and may prove useful in our studies (Kasthuri et al., 2015). Because this algorithm has already been trained and shown to effectively segment cells, it may be able to accomplish the goals we have outlined for our algorithm with the use of some modifications.

Future Directions

Additional Classification of EM data

Our current data focuses only on cells located within the S1 region of mouse cortical tissue. By similarly analyzing other brain regions, we hope to gain a whole-brain understanding of how neurons and glia differ in size and nuclear content. Furthermore, additional classification of other S1 datasets will be useful in confirming our results and reducing the overlap in distributions seen for our measurements. By building a larger database of manually classified EM data, we can begin to build a substantial training set for an algorithm capable of automatically classifying cell types.

Segmentation of Nuclei in X-ray dataset

A secondary goal of this study is to determine whether analyses of X-ray datasets produces results that are similar to our EM dataset. Using available data, we determined that the cell body distributions of neurons and glia have similar peaks in both X-ray and EM S1 datasets. However, there currently is not any data regarding the distribution of nuclei diameters for cells in X-ray volumes. Additional manual segmentation of X-ray datasets is therefore needed to confirm that our EM results are reproducible in X-ray volumes. This proves to be difficult, as X-ray images typically favor volume over resolution, so nuclei are significantly more difficult to distinguish by the human eye. Nevertheless, proving that our results are reproducible using X-ray imaging will aid in the transition from EM to X-ray segmentation, a necessary step in producing whole-brain connectomes.

Construction of Automatic Segmentation Algorithm

In addition to using membrane diameters to differentiate between neurons and glia, there are likely other quantitative variables that can aid in automatically classifying cell types. The area of a cell or nucleus, which is strongly associated with its diameter, is an obvious variable of interest and may provide distributions with less overlap. Similarly, considering the total volume of both a cell and its nucleus is a variable that is relatively easy to measure. Other variables, such as pixel gradient and grayscale value, are less obvious to the human observer but may still help minimize false classification in machine algorithms. Using weakly supervised neural networks, we may be able to discover other variables that can be used in conjunction with our results to improve the distinguishability of cell types.

Conclusion

We have laid the groundwork for our ongoing research regarding cell-type differentiation for cellular connectomics. Our initial survey of our EM dataset suggests that there is a correlation between cell type (neuron or glia) and the size of its body and nucleus. Furthermore, the results of our EM dataset appear to be in conjunction with X-ray data, at least in regard to cell bodies. Our current success has encouraged us to continue manually segment and classify cells in order to create a sizeable dataset sufficient for use as a training set for a CNN capable of distinguishing cell type based on body and nucleus diameter. It has also prompted us to further investigate the translational significance of our work to X-ray based connectomics.

Our research has multiple applications in various areas of neuroscience. By improving and expanding on connectomic data, we are becoming increasingly aware of how the structure of the brain is critical to its function. This has allowed us to make huge advancements in the medical field, where connectomes are being used to diagnose patients with a variety of disorders

(Despotović et al., 2015**)**. Moreover, our data expands on the currently existing literature on the properties of glial cells. These cells are proving to play a critical role in a variety of neurological disorders; once thought to play a passive role in the nervous system, these cells have now been liked to disorders such as multiple sclerosis, neuromylitis optica, and alzheimers (Fields, 2010). By focusing our attention on the structural nature of glia through cellular connectomics, we hope to further understand their importance in maintaining normal brain function.

References

Almad, A. A., & Maragakis, N. J. (2012). Glia: An emerging target for neurological disease therapy. *Stem Cell Research & Therapy*, 3(5), 37. doi:10.1186/scrt128

Berger A. (2002). Magnetic resonance imaging. *BMJ* (Clinical research ed.), 324(7328), 35.

Beyer, J., Hadwiger, M., Al-Awami, A., Jeong, W., Kasthuri, N., Lichtman, J. W., & Pfister, H. (2013). Exploring the Connectome: Petascale Volume Visualization of Microscopy Data Streams. *IEEE Computer Graphics and Applications*, 33(4), 50-61. doi:10.1109/MCG.2013.55

Despotović, I., Goossens, B., & Philips, W. (2015). MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Computational and Mathematical Methods in Medicine*, 2015, 1-23. doi:10.1155/2015/450341

Dyer, E. L., Gray Roncal, W., Prasad, J. A., Fernandes, H. L., Gürsoy, D., De Andrade, V., Fezzaa, K., Xiao, X., Vogelstein, J. T., Jacobsen, C., Körding, K. P., … Kasthuri, N. (2017). Quantifying Mesoscale Neuroanatomy Using X-Ray Microtomography. *eNeuro*, 4(5). doi:10.1523/ENEURO.0195-17.2017

Egmont-Petersen, M., De Riddler, D., & Handels, H. (2002). Image processing with neural networks—a review. *Pattern Recognition*, 35(10), 2279-2301. doi:10.1016/S0031-3203(01)00178-9

Fields, R. D. (2010). Central role of glia in disease research. *Neuron Glia Biology*, 6(02), 91-92. doi:10.1017/s1740925x10000177

Fieres, J., Schemmel, J., & Meier, K. (2006). Training convolutional networks of threshold neurons suited for low-power hardware implementation. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*,21-28. doi:10.1109/ijcnn.2006.246654

Gong, T., Liu, R., Tan, C. L., Farzad, N., Lee, C. K., Pang, B. C., . . . Zhang, Z. (2007). Classification of CT Brain Images of Head Trauma. *Pattern Recognition in Bioinformatics Lecture Notes in Computer Science*, 401-408. doi:10.1007/978-3-540-75286-8_38

Human Connectome Project Gallery. (n.d.). Retrieved December 8, 2018, from http://www.humanconnectomeproject.org/gallery/

Indolia, S., Goswami, A. K., Mishra, S., & Asopa, P. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science,132*, 679-688. doi:10.1016/j.procs.2018.05.069

Kasthuri, N., Hayworth, K., Berger, D., Schalek, R., Conchello, J., Knowles-Barley, S., . . . Lichtman, J. (2015). Saturated Reconstruction of a Volume of Neocortex. *Cell*, 162(3), 648-661. doi:10.1016/j.cell.2015.06.054

Mori, S., Crain, B. J., Chacko, V. P., & Zijl, P. C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology, 45(2), 265-269. doi:10.1002/1531-8249(199902)45:23.0.co;2-3