Comprehensive automated analysis of three-dimensional brain cell-resolution imaging data

William Donaldson

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Bioengineering

University of Washington

2015

Committee:

Eric Chudler

William Shain

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Comprehensive automated analysis of three-dimensional brain cell-resolution imaging data

William Donaldson

Chair of the Supervisory Committee:
Dr. Eric Chudler
Bioengineering

Neuroimaging techniques provide immensely valuable data for medical professionals and academic researchers alike, but analysis of these data is difficult and often relies on subjective interpretations. This project proposes and implements an automated method for comprehensive analysis of brain cell-resolution images acquired through fluorescent microscopy techniques. The image data are prepared using digital signal processing techniques. Next the images are automatically processed in FARSIGHT, an image processing toolkit, to segment, classify, and finally trace microglia. The robustness and viability of this process is shown in its application to several different research projects. This method improves on current analytical techniques with increases in speed, degree of automation, quantifiability, robustness, and repeatability. Automated FARSIGHT-based analysis such as this technique can enable advance research inquiries in the field of neuroscience that were previously inaccessible due to insufficient analytical capabilities.

Comprehensive automated analysis of three-dimensional brain cell-resolution imaging data

Author:
William DONALDSON

Supervisor:
Dr. William SHAIN

A thesis submitted in fulfillment of the requirements
for the degree of Master of Science

in the

Shain Research Group
Department of Bioengineering

August 2015

Declaration of Authorship

I, William DONALDSON, declare that this thesis titled, 'Comprehensive automated analysis of three-dimensional brain cell-resolution imaging data' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

"The reason most people do not recognize an opportunity when they meet it is because it usually goes around wearing overalls and looking like Hard Work."

-Henry Dodd

"I like nonsense; it wakes up the brain cells."

-Dr. Suess

UNIVERSITY OF WASHINGTON

Abstract

William Shain

Department of Bioengineering

Master of Science

Comprehensive automated analysis of three-dimensional brain cell-resolution imaging data

by William Donaldson

Neuroimaging techniques provide immensely valuable data for medical professionals and academic researchers alike, but analysis of these data is difficult and often relies on subjective interpretations. This project proposes and implements an automated method for comprehensive analysis of brain cell-resolution images acquired through fluorescent microscopy techniques. The image data are prepared using digital signal processing techniques. Next the images are automatically processed in FARSIGHT, an image processing toolkit, to segment, classify, and finally trace microglia. The robustness and viability of this process is shown in its application to several different research projects. This method improves on current analytical techniques with increases in speed, degree of automation, quantifiability, robustness, and repeatability. Automated FARSIGHT-based analysis such as this technique can enable advance research inquiries in the field of neuroscience that were previously unaccessible due to insufficient analytical capabilities.

Acknowledgements

Contents

List of Figures

Dedicated to my friends and family, who have always been there for me...

Chapter 1

Introduction

1.1 Background and Significance

The purpose of this paper is to present a method for analyzing neuroimaging data. As our ability to capture images of the brain has improved over the decades, our ability to analyze these images in a coherent fashion has remained somewhat constant. Especially with the focus of research moving to cellular and subcellular levels, the complexity of the brain remains staggering. The human brain contains hundreds of billions of neurons and other cells, forming hundreds of trillions of synapses[1], and resulting in a network so intricate as to be impossible for any given researcher to fully comprehend. This staggering complexity provides the motivation for analytical tools that can help us answer real and pressing questions in neuroscience.

With this motivation I have developed a comprehensive, automated technique for analysis of neuroimaging data collected with fluorescence microscopy. It begins and focuses heavily on the necessary steps to prepare images for optimal use in FARSIGHT and then addresses the process used in FARSIGHT itself before finally arriving at the quantified results necessary for scientific analysis. Specifically this project focuses on segmentation and classification of brain cells and the tracing of microglia.

In this first chapter, I discuss the progression of neuroimaging techniques as well as the difficulties and limitations of analyzing them. I will then describe the goals of this project. In the second chapter I discuss the basic techniques which are necessary to collect the images I have analyzed. In the third chapter I explain the metrics and techniques used for image analysis in this project. The fourth chapter describes the final analytical process from start to finish. The fifth chapter looks at examples of this process and demonstrates its robust application to various images. In the sixth chapter I discuss

the limitations of this project and suggest directions for future work. The seventh and final chapter serves as a conclusion and discussion of this project as a whole.

The tools described in the paper are based on the FARSIGHT toolkit, developed by the Roysam Lab at the University of Houston[2]. Firstly, it is important to recognize that FARSIGHT is a toolkit, not a polished software package with a unified graphical user interface. Rather, FARSIGHT is a collection of software modules which each handle tasks such as image data handling, image segmentation, feature tracing, editing or inspecting, and various analytical algorithms. While it provides the foundation of an extremely powerful tool it is necessary to link these modules and automate the processes to solve image analysis problems. Fortunately, this is a doable task because FARSIGHT is an open-source software project. Development of FARSIGHT to meet a specific need is not only accessible, but encouraged. FARSIGHT is designed primarily to analyze images generated from fluorescence microscopy and this project focuses on this modality as well. Fluorescence microscopy generates complex and voluminous data sets and FARSIGHT aims to extract meaningful measurements to quantify and analyze these data. The large data sets of optical microscopy and the complex algorithms of FARSIGHT present the challenge of large-scale computations and accentuates the limitations of modern computer systems. While only limited parallelization is implemented currently, it is part of the ultimate goal of the FARSIGHT project to be able to fully distribute the problem solving techniques over a computer network. Also incorporated into this project is a program called FIJI, which is an open-source image processing package based on ImageJ[3]. It has been chosen for its ease of use and ability to rapidly prepare images for analysis with FARSIGHT.

It is important to understand both the historical developments of neuroimaging as well as the difficulties of image analysis to better grasp the significance of this method.

1.1.1   Neuroimaging History

The beginnings of neuroimaging are not necessarily obvious. Relevant to the methodology presented in this paper was Camillo Golgi's silver staining technique developed in 1873[4]. Originally called the black reaction (*la reazione nera*), his technique involved impregnating cells with potassium dichromate and silver nitrate to allow fixed nervous tissue to be easily visualized under light microscopy. This technique required fixed tissue, meaning it could not be used on live subjects, though it provided great detail of structural information. Arguably the first non-invasive technique however was developed by 19th century Italian physician Angelo Mosso[5]. His invention, known as the human

circulation balance, was able to measure the redistribution of blood during neural activity. Unfortunately for Mosso, his work was only briefly described by William James in 1890 and went largely unacknowledged and unknown until the rediscovery of his original manuscripts by Stefano Sandrone and colleagues in 2013[6].

The next major developments in neuroimaging came from Walter Dandy with his development of ventriculography in 1918 and pneumoencephalography (PEG) in 1919[7]. Then in 1927, Egas Moniz developed cerebral angiography, allowing blood vessels to be imaged with even greater precision[8].

The 1970s heralded more advanced non-invasive imaging techniques, highlighted by Cormack and Hounsfield's computerized axial tomography (CAT or CT)[9]. Structural images of the brain became more and more detailed. By the 1980s, the development of radioligands allowed single photon emission computed tomography (SPECT) and positron emission tomography (PET) scanning of the brain as well[10]. Concurrently developed by Peter Mansfield and Paul Lauterbur, magnetic resonance imaging (MRI) would come to play a major role in neuroimaging as well[11]. The blood flow changes that could be detected by PET could also be measured with function magnetic resonance imaging (fMRI) but without radiation exposure and lower invasiveness and by the 1990s fMRI dominated the field of neural mapping[12].

(a) CT Scan of Brain[13]   (b) MRI Scan of Brain[14]   (c) PET Scan of Brain[15]
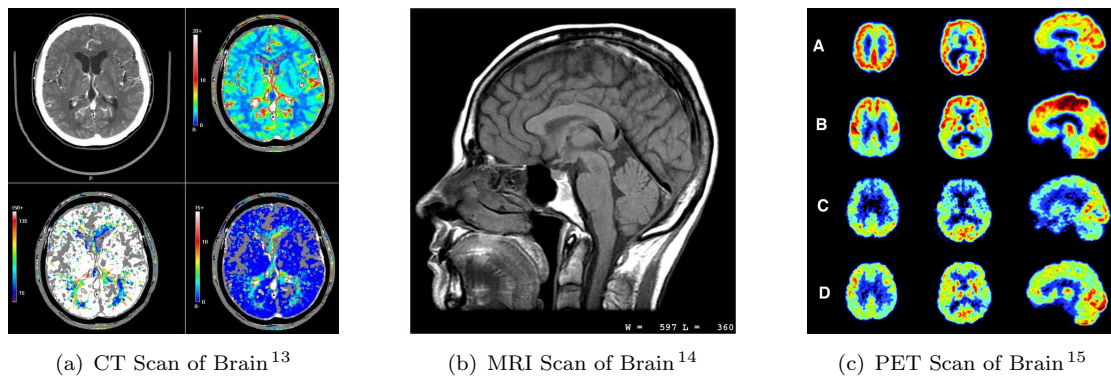
FIGURE 1.1: Comparison of images from different neuroimaging techniques.

All of these techniques improved on the state-of-the-art in some fashion, be it with increased sensitivity, better resolution, or lower invasiveness. **However, they all necessitate a subjective analysis of the data.**

1.1.2 Image Analysis

Neuroimaging techniques have improved greatly in quality, but the ability to analyze the resultant images has not kept pace in advancement. All of the techniques discussed so

far require some element of subjective analysis and interpretation of the data. Results of a CT or MRI scan require a highly trained professional to interpret, and come with error rates of up to 30%, largely based on inexperience and personal variances of the diagnostician[16].

Current qualitative techniques are far from perfect as well. Stereology for example attempts to recreate a model of a 3-dimensional space based on a statistical sampling from 2-dimensional images[17]. This technique depends explicitly on the fundamentals of geometry and statistics. This makes it very useful for interpreting well-defined systems, but the underlying assumptions leave it inadequate for hypothesis-driven studies of cell-resolution data. We cannot feasibly discover new patterns nor adequately understand complex networks of cells when those data are generated based on a statistical interpretation of a small sample size.

So for cell-resolution images, that leaves only hand-counting as a qualitative method of analysis. While it does not rely on a statistical assumption, it is obvious that hand-counting is not a viable long-term solution to image analysis. Our current laboratory setup collects up to 400 slices for each of up to 30 or more fields for a given sample, measuring 100s of GB in size. This would result in exceedingly many hours of counting cells by hand - a task both tedious and prone to mistakes. Thus, there is great need for automated analytical techniques for cell-resolution brain images based on quantitative, reproducible measurements.

The field of computed (digital) image analysis is nearly as old as digital images themselves. Application of digital geometry, pattern recognition, and signal processing algorithms has been used to solve image analysis problems in many fields and of many complexities. Tasks as simple as scanning a bar code or as complex as facial recognition in a crowd involve digital image analysis, often with quantitative, reproducible results[18]. Application of these techniques to FARSIGHT and specifically to the preparation of images for use with FARSIGHT is a promising approach that is explored in this project. Digital signal processing techniques are especially potent for generating quantitative, reproducible results to improving image quality in preparation for analytical image manipulation.

This is where FARSIGHT comes in. FARSIGHT contains quantitative tools suitable for studying complex and dynamic biological systems imaged with 4D and 5D microscopy techniques. Complex neuroscientific inquiries can be made using the analytical techniques in the FARSIGHT toolkit. The actual process to get to this stage is not trivial however, and this paper addresses the necessary steps to working with FARSIGHT to solve complex image analysis problems.

1.2   Goals of this Project

The goal of this project is to enhance the experience of a researcher who desires a comprehensive, quantitative analysis of cell-resolution imaging data using the FARSIGHT toolkit.

Development of FARSIGHT has resulted in exceptionally powerful algorithms for answering many types of questions relevant to neuroscience. Enabling their use by researchers is still a necessary step however. Regardless of the power of a tool, knowledge of its appropriate use is essential to its success. This paper begins this process of enabling by developing and documenting a largely-automated process for use of FARSIGHT in some neuroscience applications.

By analyzing the needs of researchers, I have identified three key elements necessary to the success of this project: automation, efficiency, and accessibility.

1.2.1   Automation

The process described within should be highly automated to reduce the number of necessary man-hours devoted to processing and allow more time for critical analysis of the data. Using the tools provided in FARSIGHT typically requires close attention by the researcher to initiate and follow through each step in the process. This is not only time consuming but extremely tedious and not resistant to mistakes or lapses in consistency. Automation of the process will ensure consistent processing of all the data, prevent mistakes, and reduce the amount of time that a researcher must directly interact with the software.

1.2.2   Efficiency

While FARSIGHT can be inherently powerful it is not inherently efficient. In order for the end result to be a desirable tool it must be efficient in terms of both researcher man-hours and the computation power required. Current techniques require samples be processed in serial. This project seeks to implement a batch algorithm to ensure consistency and further reduce the involvement of individual researchers. Eventually, parallelization of this process should be implemented in order further reduce the time requirement and allow the computation power to be distributed over a network of computers. Unfortunately that goes beyond the scope of this particular project.

1.2.3 Accessibility

The FARSIGHT toolkit provides powerful algorithms but their use was not well documented. Regardless of the strength of a technique, it must be easily approachable by the desired user in order to be adopted in practice. The field of neuroscience contains researchers of many different skill sets; from undergraduates to heads of departments and all levels in between. Thus we can assume our target audience to be intelligent, but it would be unwise to assume advanced knowledge of computer systems or even the basics of command line interaction with a computer. In order to be the most effective tool, FARSIGHT must be easily usable and this means the language used in documentation must be precise but clear and that any actions necessary to complete the process must be simple and straightforward.

As additional requirement for accessibility, the process must be robust and yield quantifiable results that can be analyzed appropriately. Many types of projects should be able to use this process. The techniques developed must be robust to many types of variables, including the species of animal, scale of analysis, and disease status. The outputs must have some level of quantifiability as well in order to achieve reproducible analysis and be useful in a scientific setting. These two factors, while not limiting the accessibility directly, impact the desire of a researcher to use FARSIGHT as a tool. If FARSIGHT does not work easily to solve a given problem, nor generate consistent quantifiable outputs then it is as if the researcher did not have access to the tools that FARSIGHT provides.

Chapter 2

Methods

2.1 Beginning Work

Experimental setup is an essential part of this process. While the purpose of FARSIGHT is to enable all manner of neuroscientific inquiry, this chapter will provide one example of the methods used to collect data in preparation for analysis with FARSIGHT as well as introduce the digital processing that follows. The techniques documented are for an *in vivo* rat model and the results are documented in Chapter 5.

The first step is the surgical preparation of the specimen, followed by immunohistochemistry to label the tissue. After that the tissue is imaged using confocal miscroscopy and the data are stored on the server for computer analysis. The data are then prepared in a large batch in FIJI for processing in FARSIGHT. Once in FARSIGHT, a training set is generated for cell classification and then all images in the set are segmented and classified based on that training data. Lastly, the glial cells are traced and quantifiable measurements are generated for the researcher.

2.1.1 Surgical Approach

Surgeries were performed in two phases: implantation and perfusion. The purpose of the first surgical phase was the implantation of the electrode array into the cortex of the rat. Rats were anesthetized with isoflurane. The head of the rat was shaved and cleaned with betadine to prepare a sterile surgical site for incision. The head was opened to reveal the bregma skull landmark and surgical screws were implanted to provide a firm attachment point for the polymethyl methacrylate (pMMA) skull cap later. A dremel was used to cut a roughly 2 mm square window to the right hemisphere. The dura was carefully cut to avoid damaging the cortex of the brain before the device was

inserted. Then a single-shank electrode device was inserted to a depth of 1 mm with the pneumatic inserter. After the device was inserted, a pMMA cap was applied over the device to protect it from the animal and prevent it from becoming dislodged, and then the animal was closed.

The animal was left for a nominal time of 5 minutes after implant before the perfusion began. This was to ensure that the data accurately represented only the very early initial impact effects of the device, while also providing a consistent timeline from insertion to termination of the animal. The purpose of the perfusion was to terminate the animal and fix the brain immediately to prevent any degradation of the tissue that would occur *post mortem*. The rat was anesthetized with isoflurane and placed on ice to cool the body. An incision was made under the ribcage in order to expose the thoracic cavity. Then an incision was made in the right atrium and left ventricle of the heart and a cannula was inserted into the left ventricular incision. The animal was perfused with buffer at a constant pressure of 100 mmHG for exsanguination and then immediately with 300 mL 0.4% paraformaldehyde (PFA) at a constant pressure of 60 mmHg to fix the tissue. The head was decapitated with a surgical guillotine and the brain was dissected out. The dissected brain was stored at 20C in PFA for 24 hours and then transferred to PBS to prevent further crosslinking and degradation.

2.1.2   Immunohistochemistry and Imaging

Immunohistochemistry was performed over a three day period. On Day 1 brain tissue was sliced in 300 micron thick sections so as to guarantee some sections would avoid the shaft of the electrode. A 4% agar backing was used for stabilization and cleaner sectioning results. After sectioning, samples were transferred sequentially to 24-well plates in HBHS w/ sodium azide. Samples were washed three times in the HBHS w/ azide. The samples were then incubated at ambient temperature in a 5mg/ml solution of NaBH4 for 30 minutes to expose the antigen after crosslinking. The 3x buffer wash step was repeated and the samples were incubated at ambient temperature in a 0.2% Triton-X-100 solution for 30 minutes. The Triton-X-100 solution is a detergent and lyses the membrane. The samples were once again washed three times with buffer and then incubated in 0.3% Hydrogen Peroxide (H_2O_2) at room temperature for 30 minutes in order to block endogenous peroxidases. The buffer wash step was repeated again. The samples were incubated at ambient temperature in Image-iT FX for 30 minutes to block background staining and non-specific dye interactions. The samples were washed one last time and incubated overnight at ambient temperature in a 0.2% Triton-X-100 solution with the primary antibodies (Rabbit anti-Iba1 for Microglia).

Day 2 of immunohistochemistry began with four 30-minute washes in the HBHS w/ sodium azide buffer. Then the samples were incubated overnight at ambient temperature in a 0.2% Triton-X-100 solution with the secondary antibodies (Hoechst for nuclei and anti-Rabbit 488 for Microglia). The 24-well plate was covered with foil in order to limit the exposure to light and prevent photobleaching of the organic dyes.

Day 3 of immunohistochemistry began with four 30-minute washes in HBHS w/ azide and a 0.5% Tween 20 solution. This reduces the non-specific binding and protein-to-protein interactions. The samples were then mounted on glass slides with Fluoromount G in preparation for imaging. After the samples were prepared and mounted to slides, they were imaged with 3D confocal microscopy. Scans were taken as a single field image at 30x magnification. Each cell type was isolated by the specific fluorophores of the secondary antibody.

2.2   Computer Analysis

Once the imaging data are acquired, the digital processing necessary for analysis of the data can begin. This process will be documented in great detail in Chapter 4 and Chapter 5, so this section will serve only as an introduction. The first step is preparation of the data within FIJI. This is done via a batch automation plugin that filters and subtracts background luminance to increase the signal-to-noise ratio. After this an image is loaded into the FARSIGHT Nucleus Editor and run through the segmentation algorithm. Using this segmented result, an active learning session is started to develop a classification model. This process requires both a graphic user interface and direct intervention by the researcher to identify different cell types. This step is essential to the quality of success of the analysis but only needs to be performed once per experiment as the classifier model can be applied to all data within a set. Afterwards, all images are segmented and classified based on this model using FARSIGHT. Lastly, microglia are traced and the resulting images can be viewed, manipulated, and analyzed within the FARSIGHT Trace Editor.

Chapter 3

Signal Processing

3.1 Signal-to-Noise Ratio

An important concept in this paper is signal-to-noise ratio (SNR), which compares the amount of a desired signal to the amount of background noise. SNR is defined as the ratio of the power of the signal to the power of the noise and can be expressed in both ratio form or in decibels[19]. The industry standard for image analysis typically expresses SNR in decibels, but where possible I will include both a decibel and ratio value.

SNR is an extremely useful measurement for this process. The images collected are too complex to easily tell if one image is better than another. For example, take the images in Figure 3.1. At first glance it is difficult to tell how these images differ but the SNR gives us a quantifiable measurement to compare them. I will demonstrate later the effects of an increased SNR on the FARSIGHT results.

Ideally, images in a data set supplied to FARSIGHT should have an exceptionally high SNR. The Rose Criterion states that to be 100% certain that the data are real the SNR should be at least 5.0 or approximately 14 dB[20]. For photographic images, the ISO standards holds that acceptable image quality should have an SNR of 20 dB and ideally of up to 32.04 dB[21]. As can be seen in the images in this chapter, this method results in images approaching the Rose Criterion. As development of FARSIGHT continues and methods of data collection improve it is not unlikely that images will reach the 20 to 30 dB range of SNR.

(a) SNR = 11.5 dB

(b) SNR = 12.0 dB

(c) SNR = 12.8 dB

(d) SNR = 13.2 dB
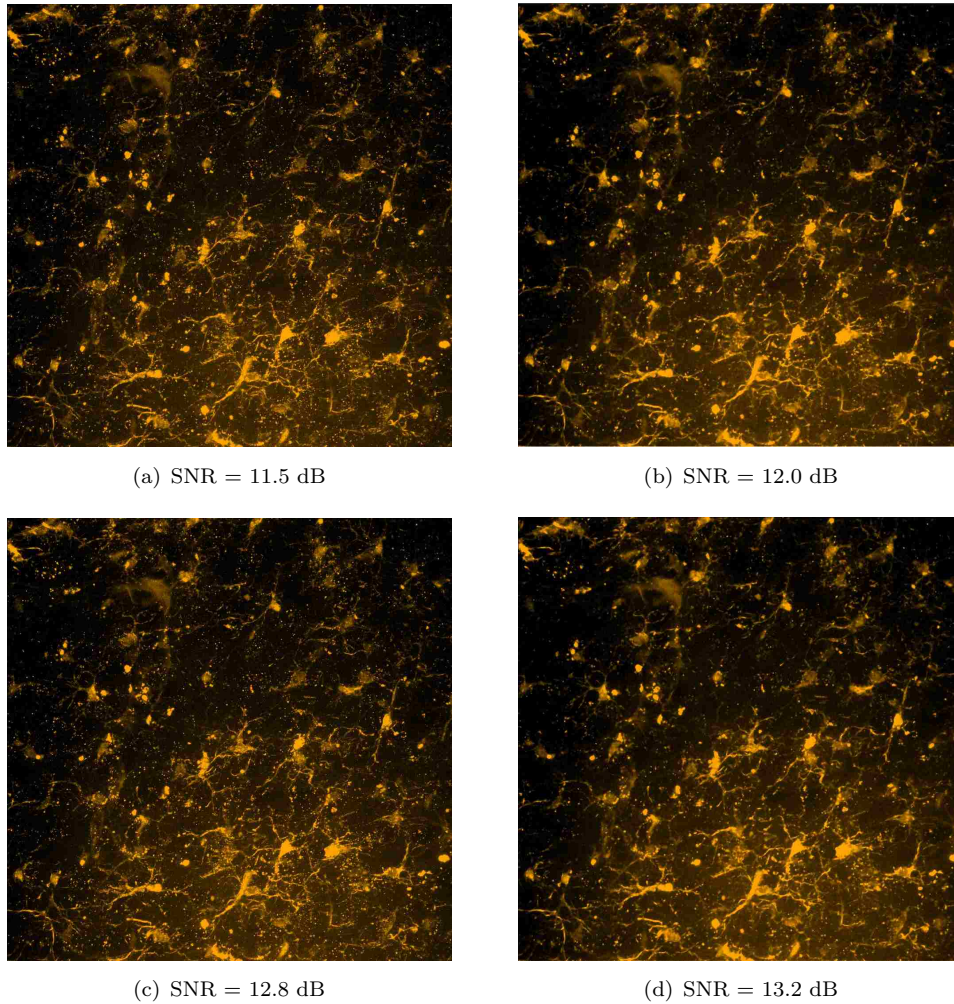
FIGURE 3.1: Comparison of the same image after several stages of preparation. Note how similar these images appear at first glance, though the SNRs differ by almost 15%.

3.1.1 Measuring Signal-to-Noise Ratio

Signal-to-noise ratio is defined as the ratio of the power of the signal to the power of the noise[22], which is mathematically expressed as:

$$SNR = P_{signal}/P_{noise}$$

While this is a precise mathematical definition it is difficult to use for the purpose of comparing image processing techniques because of the difficulty in differentiating signal from all noise and calculating an average power for the two quantities. Thus, an alternative definition is commonly used for image processing applications[19]. This alternate definition is given by:

$$SNR = \mu_{signal}/\sigma_{background}$$

In this definition μ represents an expected value and σ represents the standard deviation of the noise. Note that it may be difficult or impossible to know the exact standard deviation of the noise so an estimate is used. While this definition does not have the exactness associated it is much more readily calculable given image processing resources. Specifically, μ is calculated as the average pixel value and σ is calculated as the standard deviation of the pixel values in a given (background) region. Note also that this definition is only usable when the values are always non-negative, which is in fact the case for the pixel values in these images[22].

Under this definition there are natural limitations of image files to be aware of. Given FARSIGHTs use of fluorescent microscopy the peak foreground signal is white and thus the background is defined as black. In order to maximize the ability of FARSIGHT to detect the foreground signal, the background of images often becomes 'clamped' to pure black (a hexadecimal RGB value of 000000). Thus if all or most of the background is clamped to a specific zero-value, the standard deviation of the background tends to zero and the SNR tends toward infinity[19]. While improving the SNR is valuable to this project, artificially boosting it severely limits its use as a comparative measurement. To combat this we can use the approximation that the standard deviation of the signal is equal to the standard deviation of the noise[19]:

$$\sigma_{background} \approx \sigma_{signal}$$

$$\therefore SNR \approx \mu_{signal}/\sigma_{signal}$$

This definition of SNR is perfectly serviceable for many image processing applications. However, it is still hindered as a true comparative tool because the size of the area used to calculate σ_{signal} is not implicit in the definition. It is possible to find a quantitative measurement of noise from only the data in the image by calculating root-mean-square noise (RMS noise)[23]:

$$RMS_{noise} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \frac{\sum_{i=1}^{n}X_i}{n}\right)^2}{n}}$$

Where n is the number of lines in the image and X_i is the value of ith line in the image. This calculation uses only values taken directly from an image and is not sensitive to different sized regions. Combining the standard logarithmic transformation used for signal processing and this final measurement allows us to calculate the SNR of images

for comparative purposes with the following equation:

$$SNR = 20 \log_{10} \left(\frac{\mu_{signal}}{RMS_{noise}} \right) dB = 20 \log_{10} \left(\frac{\mu_{signal}}{\sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \frac{\sum_{i=1}^{n} X_i}{n} \right)^2}{n}}} \right) dB$$

For practical purposes, this measurement of SNR has been implemented in a FIJI plugin as well.

3.2 Image Analysis Techniques

In this section I will discuss the image analysis techniques used in this paper. These techniques are used to improve the SNR and prepare the images for analysis with FAR-SIGHT. To understand the improvements that are made, first look at a representative image before preparation.
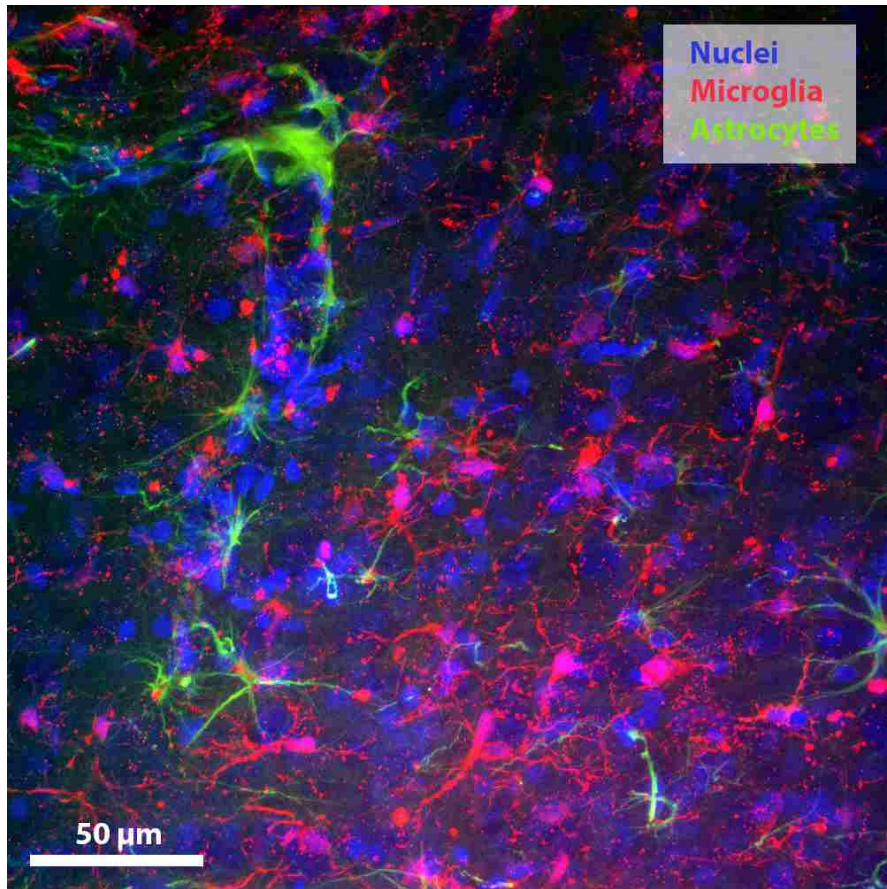

FIGURE 3.2: A representative image of rat cortex containing nuclei, astrocytes, and microglia.

Figure 3.2 shows an overlay of the nuclei, astrocyte, and microglia channels of an image. Notice that there is both high frequency noise induced by the limitations of the optical microscopy techniques and low frequency background luminance that is contributed by out of focus information. For denoising and image such as this, I have identified several important factors. The techniques applied must be fast, as one of the major goals of the project is to reduce the overall time of processing. The techniques need to address the high-frequency noise in the image as well as the low frequency background luminance. Any image processing techniques must be able to maintain the integrity of edges within the image and the connectedness of the image. If in the process of filtering out high frequency noise for example, the technique severs a connected portion of a glial process then that cell will have incomplete tracing information. As a rule, the preparation should never reduce the effectiveness of the FARSIGHT process. Lastly, the techniques used do need to assume nonlinearity and space invariance. Since the image data is stored in a file with finite color definitions (000000 for black and ffffff for white) the limits of the image are not linear; while it can be reasonable to ignore these corner cases in some circumstances, FARSIGHT needs to interact with the background and foreground so ensuring that the preparation handles these effects properly is important. Additionally, the preparation may not always be performed starting in the same location of the image data, thus a space invariant process should be used.

3.2.1   Filtering

Digital filtering is the primary method used for denoising an image. In order to improve the SNR we must either increase the signal or decrease the noise. By the time the data is collected and ready for preprocessing, the amount of signal is fixed so our best option for improving the SNR is to reduce the noise.

The filter used is a median filter, which runs through the signal entry-by-entry and replaces the value with the median of the neighboring points. The median filter is computationally simple (and therefore fast), nonlinear, and space invariant, thus fulfilling those aspects of the requirements. It is very useful for removing the high frequency noise while also maintaining edges within the image and connectedness because high frequency outliers can be normalized without affecting the pixels around them while edges within an image provide redundancy and increase the chance that the edge value remains the median[24].

There are limitations to be considered for the median filter as well. Edges, and especially the corner of the image do not have the requisite number of pixels around them to adequately determine the median[24]. Often, it is sufficient to leave only the edge pixels
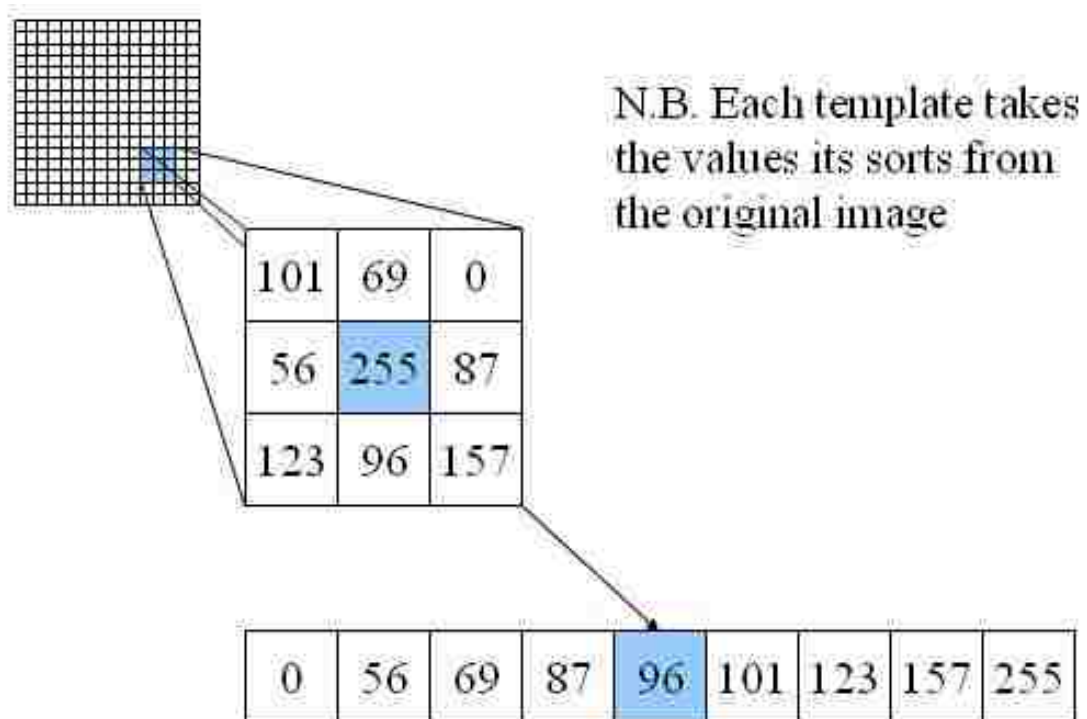
FIGURE 3.3: A visual representation of the median filter[25].

unfiltered[24]. However, in this case leaving the edges unfiltered can have unintended consequences during the tracing step as can be seen in Figure 3.4.

The value of the edge pixel was significant enough for the tracing algorithm to include it as part of a glial process, when in fact that pixel is not part of that cell at all. To address this issue for pixels on the edge of the image, the median filter extends the values of the pixels surrounding the pixel in question off the edge of the screen to create the set from which to determine the median[24]. The improved result can be seen in Figure 3.5. This edge accommodation technique is not a perfect solution, but maintains the speed of the filter and improves the result significantly enough for this purpose.
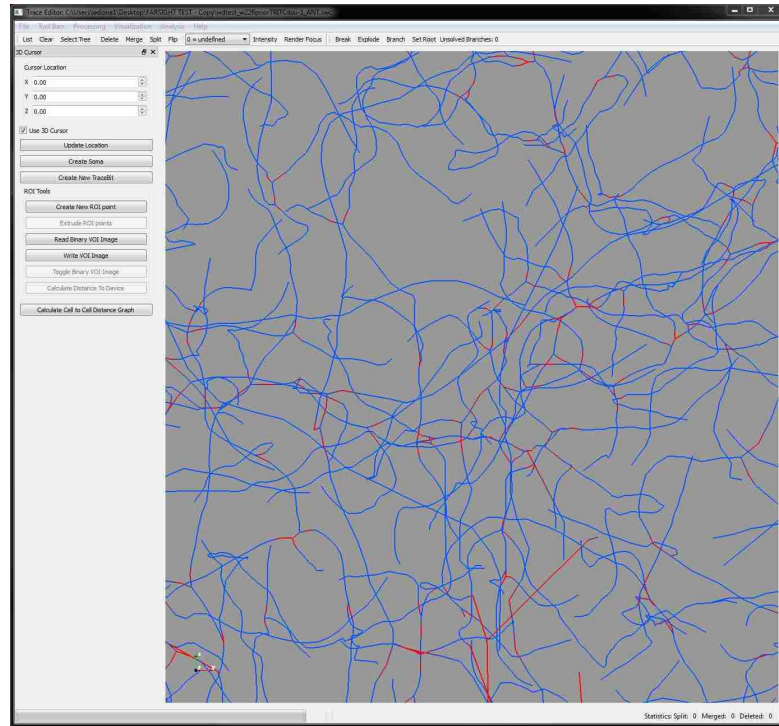
FIGURE 3.4: An example of poor tracing results. In this image the edges remained unfiltered and as a result the tracing has many incorrect terminations of processes at these edges. Additionally, the conflicts of the tracing are also shown in red where the tracing results were conflicted based on the improper preparation.

FIGURE 3.5: Corner of a tracing with proper filtering. The red circles indicate regions that have a direct improvement as a result of filtering.

An alternative technique that could be implemented is 3-dimensional median filtering. This is the same as median filtering in 2-dimensions but uses a cubic space around the pixel to determine the median. This approach can increase the SNR even further, but due to the fact th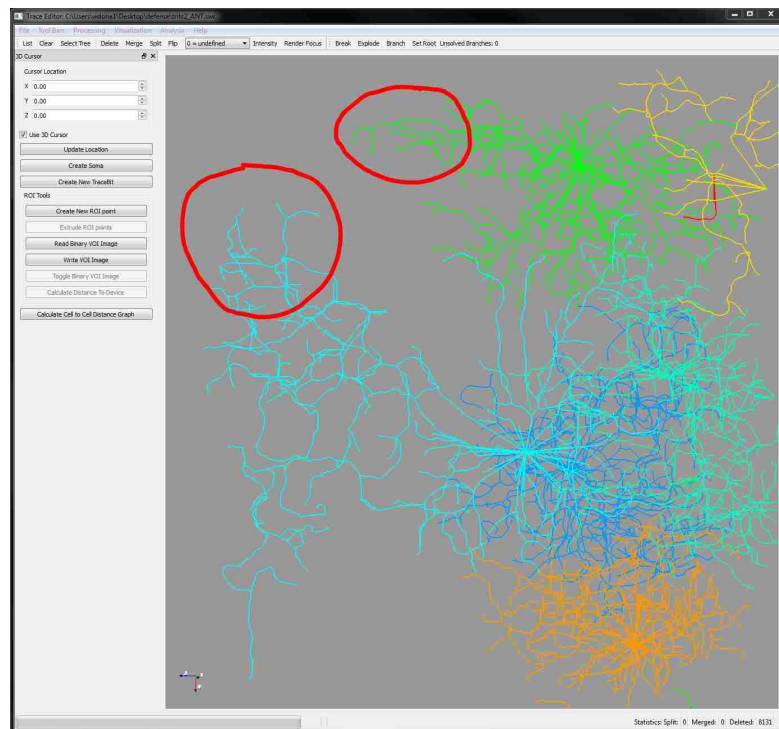at glial processes transect planes any angle it reduces the data value redundancy necessary to maintain the connectedness of the process and has a much higher risk of negatively affecting the glial tracing. Thus, standard 2-dimensional median filtering is preferred.

3.2.2   Background Subtraction

While the median filtration step handles the high frequency noise, it is still necessary to address the low-frequency luminance noise in the image. For this, the best technique is rolling-ball background subtraction. Background subtraction (equivalent to foreground detection) is a category of image processing techniques where the foreground of an image is isolated from the background for further processing. It is often used for detecting moving objects in a video stream, as logic suggests that the non-moving background image will remain somewhat static so subtracting that as a model from each frame will result in only the moving object in question[26]. The basic assumption that the background is static holds true for these optical microscopy techniques as well, and allows us to use these techniques to isolate the background and reduce the luminance. This sets a consistent floor for the pixel values in the image. Figure 3.6 shows how the low frequency, low intensity background of an image can appear before subtraction.

(a) DAPI Channel   (b) TRITC Channel
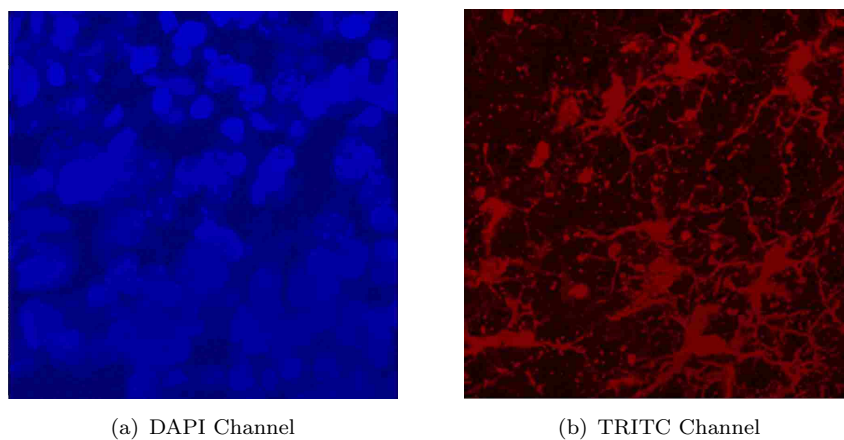
FIGURE 3.6:   Figures with exceptionally high background luminance. Both images could benefit greatly from background subtraction.

The rolling-ball algorithm is especially good at removing smooth, continuous backgrounds. First developed by Stanley Sternberg, the process can be described by imagining the image as a 3-dimensional surface with the pixel values corresponding to the height of the surface then a ball rolling over the back of the surface to level the background[27].
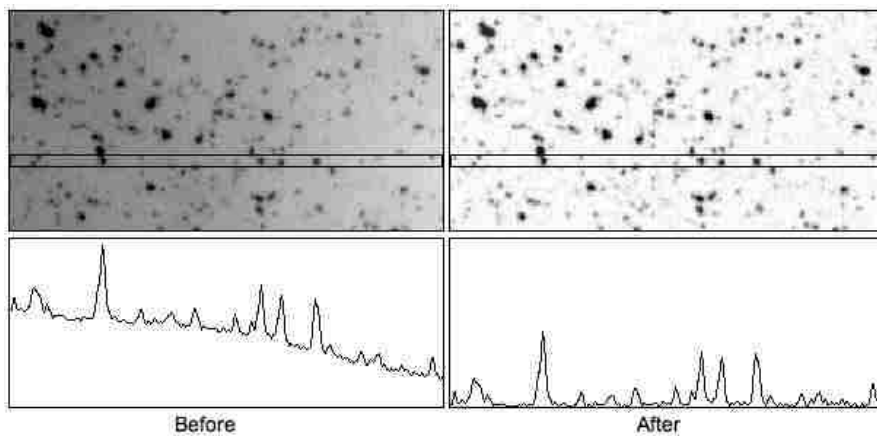
FIGURE 3.7: A visual representation of rolling-ball background subtraction[28].

The primary input to this algorithm is the rolling-ball radius, which is the radius of curvature of a parabaloid used as the "rolling ball". This value should be at least as large as the radius of the largest object in the foreground[27]. The results of background subtraction can be seen in Figure 3.8.
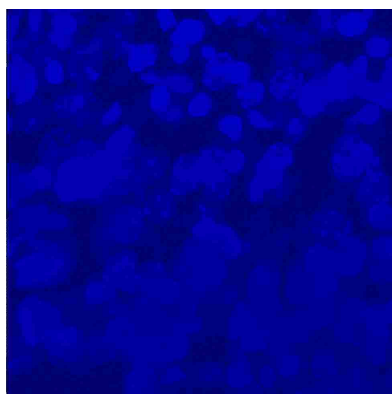
(a) DAPI Channel - Before
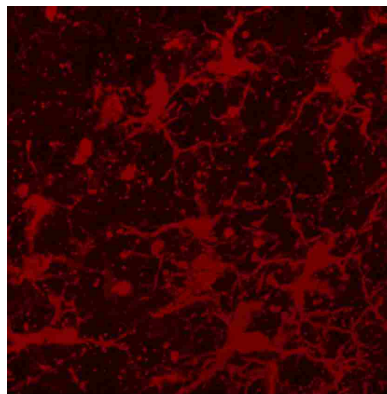
(b) TRITC Channel - Before
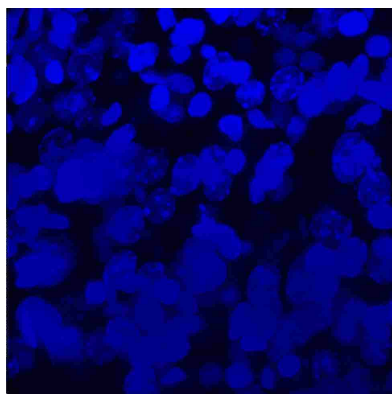
(c) DAPI Channel - After
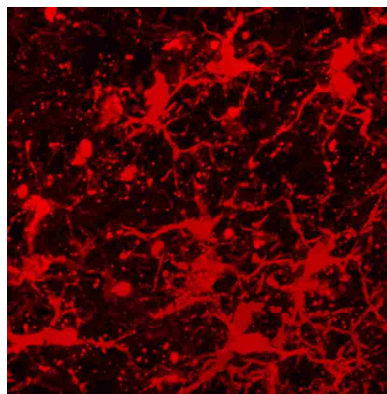
(d) TRITC Channel - After

FIGURE 3.8: Before and after images of background subtraction.

Chapter 4

Steps for Image Analysis

4.1 Overview of Digital Processing

This document provides step-by-step details for using the FARSIGHT Toolkit to analyze three-dimensional (3D), multispectral image data including nuclear segmentation, cell classification, and process tracing[2]. These techniques will enable other analytics such as cell counting and density measurements.

There are three major sections: Preparation, Process, and Analysis. Preparation explains the work necessary to prepare data for optimal use with FARSIGHT (i.e. preprocessing). Process covers the steps necessary to analyze the data using FARSIGHT and generate qualitative measurements. Analysis describes the methods to analyze the outputs of the FARSIGHT process.

4.1.1 Preparation

Proper preparation of image data is necessary for optimal performance of FARSIGHT; it reduces the time and system resources required and produces more accurate results. Preparation is performed using ImageJ or FIJI. There were four requirements for the preparation section: batch processing (or some automated process), consistency of results, robustness of application, and increase in signal-to-noise ratio. There are two phases within preparation: digital signal processing and XML file generation. The signal processing techniques chosen and the rationale for their use are described in Chapter 3. XLM files are the way that FARSIGHT interacts with multispectral images. Each image has an XML template defining its channels, and each project has two XML templates defining the complex processing done by FARSIGHT. The preparation steps are detailed below:

Preparation How to prepare image data sets for use in FARSIGHT

1: If the initial z-slices in a data set do not contain data, remove these z-slices. This will reduce the total image size and decrease the processing time

If you do remove any z − slices, then be consistent with all the other images

2: Use a median filter with a radius of 2.0 pixels to remove high frequency noise

3: Use the background subtraction algorithm with a rolling-ball radius of 50.0 pixels to reduce background luminance

These pixel values can be adjusted if necessary but are good for most projects

4: Alternatively, this process is automated via the BatchPrep plugin in FARSIGHT.

5: Create an XML file for each image in the data set that defines the channels to be used by FARSIGHT

6: Create two XML files for the entire project which define the process for FARSIGHT. The first defines the segmentation process and the second incorporates the classification steps

Templates are available for all XML files
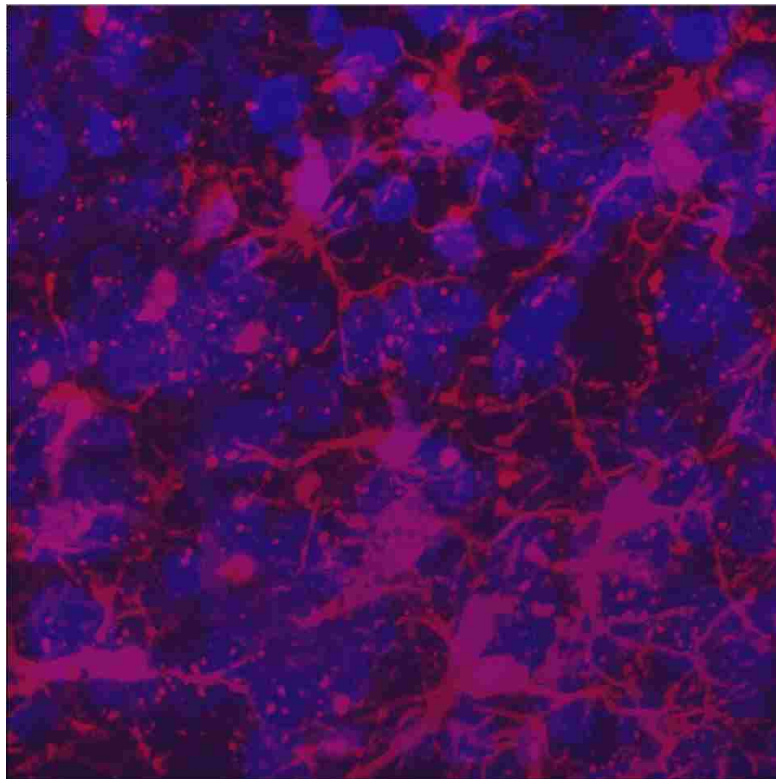
FIGURE 4.1: The basic image used for the preparation phase.

FIGURE 4.2: Resulting image after median filtering in FIJI.
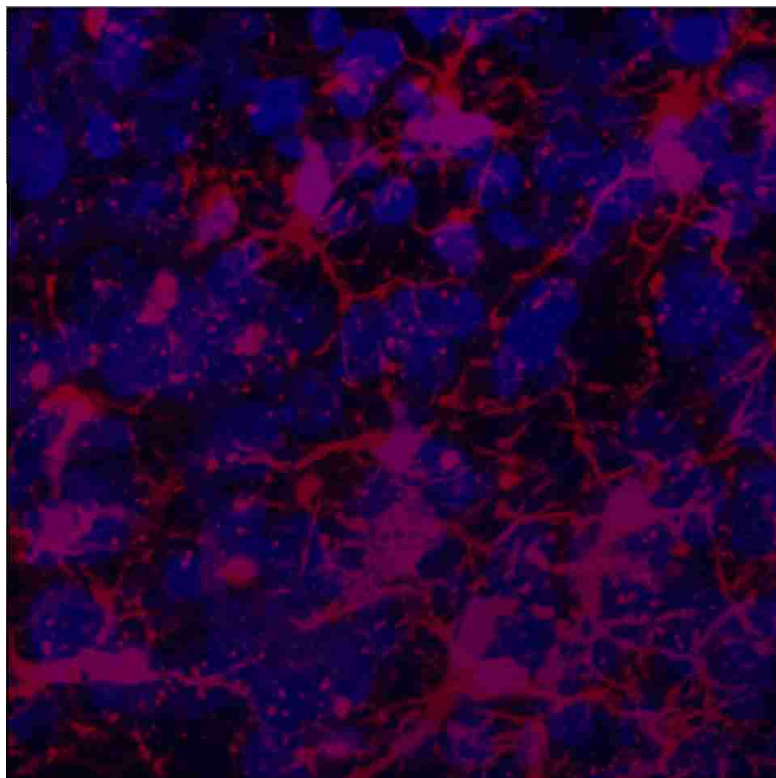
FIGURE 4.3: Resulting image after background subtraction in FIJI.

4.1.2 FARSIGHT Process

Once the data are processed in FIJI and the appropriate XML files are generated, the data are loaded into FARSIGHT for processing. This can be done either through the GUI or command line. While the command line interface is capable of processing 16-bit TIFF images, the GUI only supports 8-bit images at this time. Whe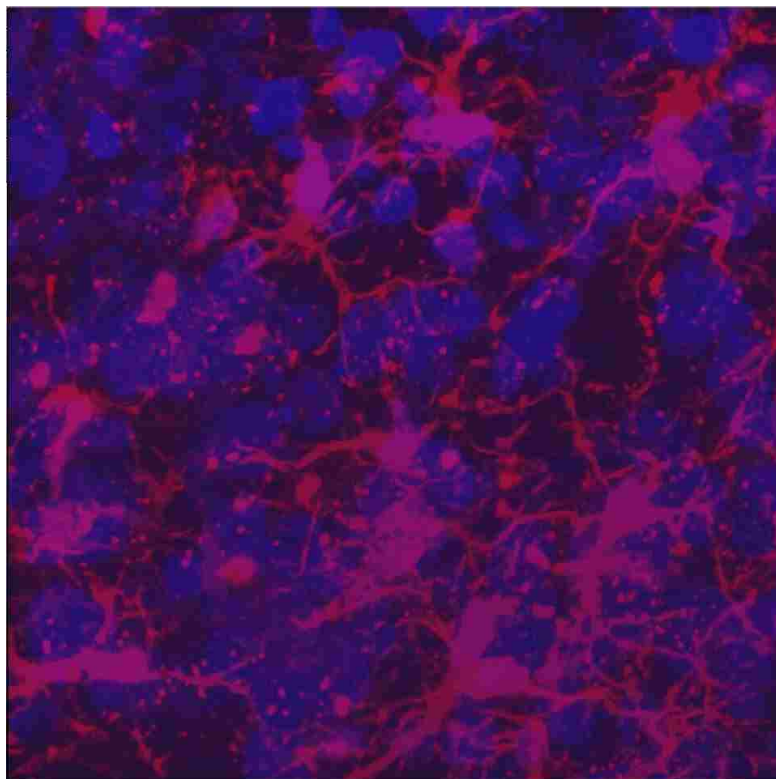n using the GUI make sure that the images have been converted to 8-bit TIFF (typical source images are 16-bit TIFF). This can easily be done in FIJI and is automatically done in the BatchPreP plugin.

The command line is the most efficient way to process these images; it is lighter on system resources and typically faster. Most of this process can be done from the command line much easier as well. There are several points where using the GUI is recommended however. It is recommended that when starting this analysis process for a new project that you open the results in the GUI after each step and verify that the results are satisfactory. Additionally when training the classifier it is necessary to use the GUI.

The FARSIGHT process is described on the next page:

Processing How to process images within FARSIGHT

1: Choose a representative image for segmentation that is likely to result in strong classification data

2: Using the command line, segment the image:

projproc.exe InputImage.xml OutputResults.tiff OutputTable.txt ProcessFile1.xml

3: Open the result in the FARSIGHT Nucelus Editor

File → Load Image, then File → Load Result

4: Identify 3 to 5 very good examples for each classification of cell. Make sure to note the IDs of each of these cells as they are listed in the table

5: Start an Active Learning Session to train the algorithm for cell classification

Select Tools → Active Learning → Start Active Learning

6: Name the training set.

7: Set the confidence threshold. This is 100/(number of channels)

If there are nuclei and microglia the threshold is $100/2\ channels\ = 50.$

8: Add classes to be consistent with the number of channels in each image

9: List the IDs of each exemplar for each class you chose. This should be a comma separated list

10: Select the appropriate features to be considered by the classification algorithm. For most projects, unselect the ID and centroids in x, y, z, and make sure that mgTOTAL and mgAVERAGE are selected at the bottom of the list

11: The active learning wizard will now ask you to identify as many cells as you can. It will present you with each classification or uncertainty as options.

12: When the wizard understand the classification problem save the classifier

Select Tools → Active Learning → Save Active Learning

Ensure: Be certain you know where you save this. You will be using this model for all the remaining images. The path where you saved this should be included in the template for segmentation with classification XML file.

13: Using the command line again, process all the images using the classification XML definition:

projproc.exe InputImage.xml OutputResults.tiff OutputTable.txt ProcessFile2.xml

14: Lastly, use the command line to trace the microglia:

MultipleNeuronTracing.exe InputFileName SomaCentroids.txt SomaImageFile options file

---

To view the resulting tracing, load it using the FARSIGHT Trace Editor.
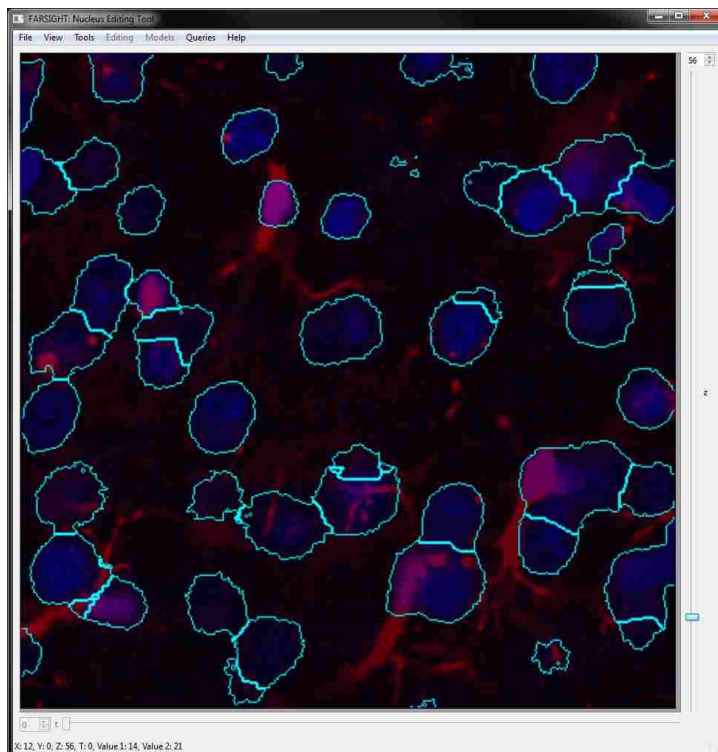
FIGURE 4.4: Results of the segmentation algorithm in FARSIGHT.
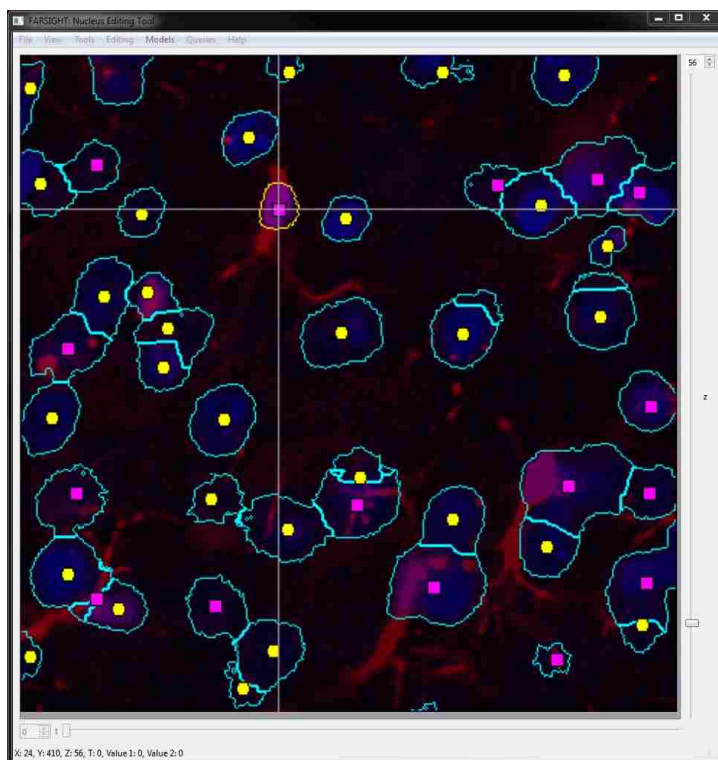

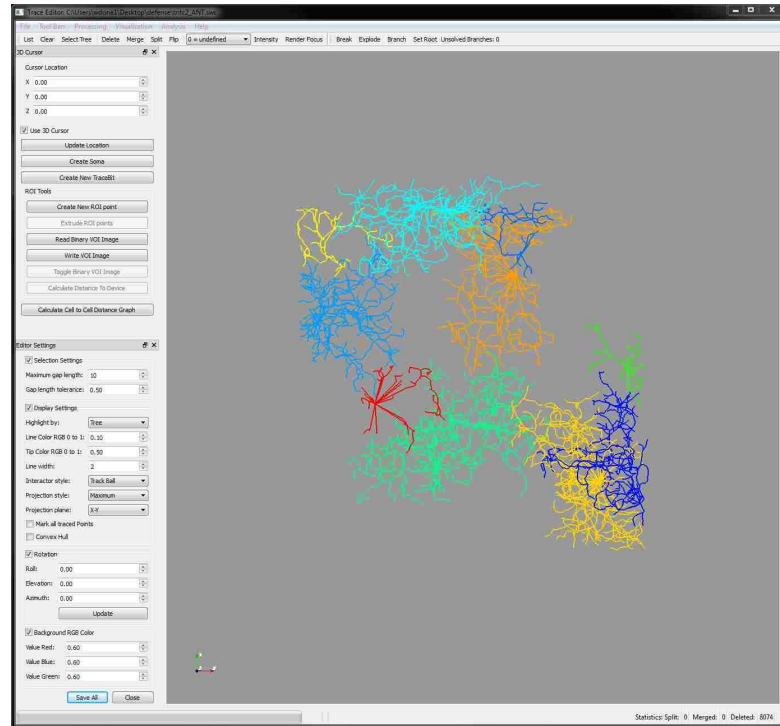FIGURE 4.5: Results of the classification algorithm in FARSIGHT
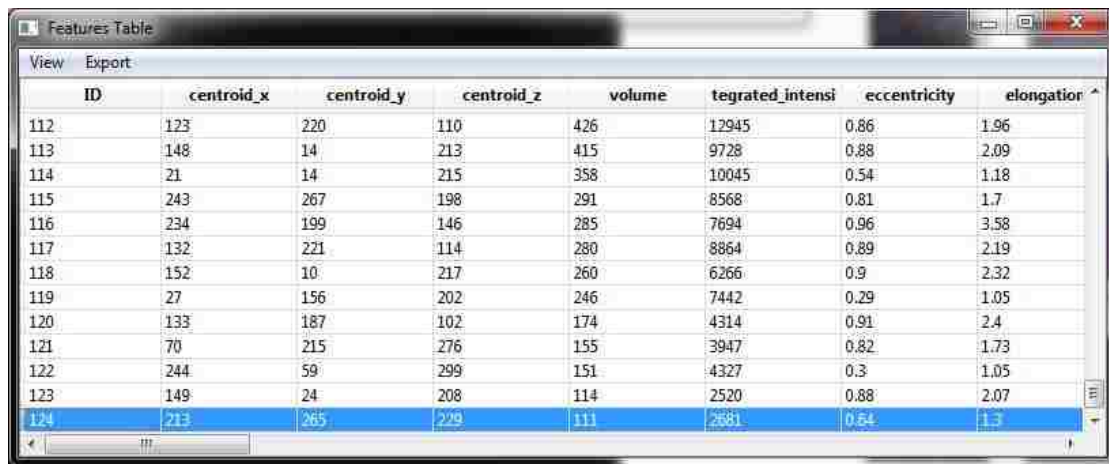
FIGURE 4.6: Results of the tracing routine in FARSIGHT.

4.1.3 Analysis

Analytical methods and desired analytes vary wildly between projects. I will describe some of the common approaches in this section, though it is far from an exhaustive list. The three analyses will be a cell count and density measurement, a cell to cell distance graph, and a Coifman harmonic co-clustering.

Cell counts are the most common analysis performed currently. This can be completed after the classification stage, or even after segmentation if the type of cell is irrelevant to the study, and does not actually require the final tracing process. Since each soma in the image is given an ID number during segmentation, the largest ID is the total number of cells in the image. The table generated in this step will also list a the classification of each cell ID, allowing for comparisons between each cell type as well. While a cell count is a very simple measurement and easy to find, it can be highly informative as well. Comparisons of cell counts around two different neuroprosthetic electrode implants could be a measurement of the biocompatibility of the device coatings[29] or the damage caused by different implant techniques[30]. Additionally, given the total number of cells and a known volume of the region they are in, it is trivially easy to calculate the cell density of that region. This could be useful for comparing an *in vitro* cell culture model to *in vivo* conditions in the brain.

ID	centroid_x	centroid_y	centroid_z	volume	tegrated_intensi	eccentricity	elongation
112	123	220	110	426	12945	0.86	1.96
113	148	14	213	415	9728	0.88	2.09
114	21	14	215	358	10045	0.54	1.18
115	243	267	198	291	8568	0.81	1.7
116	234	199	146	285	7694	0.96	3.58
117	132	221	114	280	8864	0.89	2.19
118	152	10	217	260	6266	0.9	2.32
119	27	156	202	246	7442	0.29	1.05
120	133	187	102	174	4314	0.91	2.4
121	70	215	276	155	3947	0.82	1.73
122	244	59	299	151	4327	0.3	1.05
123	149	24	208	114	2520	0.88	2.07
124	213	265	229	111	2681	0.64	1.3

FIGURE 4.7: Basic analysis of the data using a cellular count.

Cell counting is the most fundamental analysis, but only provides a limited amount of information. A more information-dense analysis requires the FARSIGHT Trace Editor. After loading the tracing results, the trace editor can perform its own cellular analytics specifically on the microglia. From there the program can generate a cell to cell distance graph. This provides the location of the microglia somas as well as their distances from each other. This provides more spatial information than a simple cell count or density measurement and can allow for conclusions based on the physical location of glial cells relative to one another.

Analysis   How to generate a Cell to Cell Distance Graph

1: Load the tracing result in the FARSIGHT Trace Editor

2: Perform the Trace Editors cellular analysis

   Select Analysis → Cell Analysis

3: Click Calculate Cell to Cell Distance Graph

Finally, Coifman harmonic co-clustering builds beyond the just the location and number of cells to generate an analytical technique based on the physical characteristics of each cell[31]. This is done by the Trace Editor as well. In one study, this was used to reveal four morphologically distinct classes of microglia that align with known activation patterns, thus enabling the researchers to map spatial distributions of microglial activation and cell abundances[31].
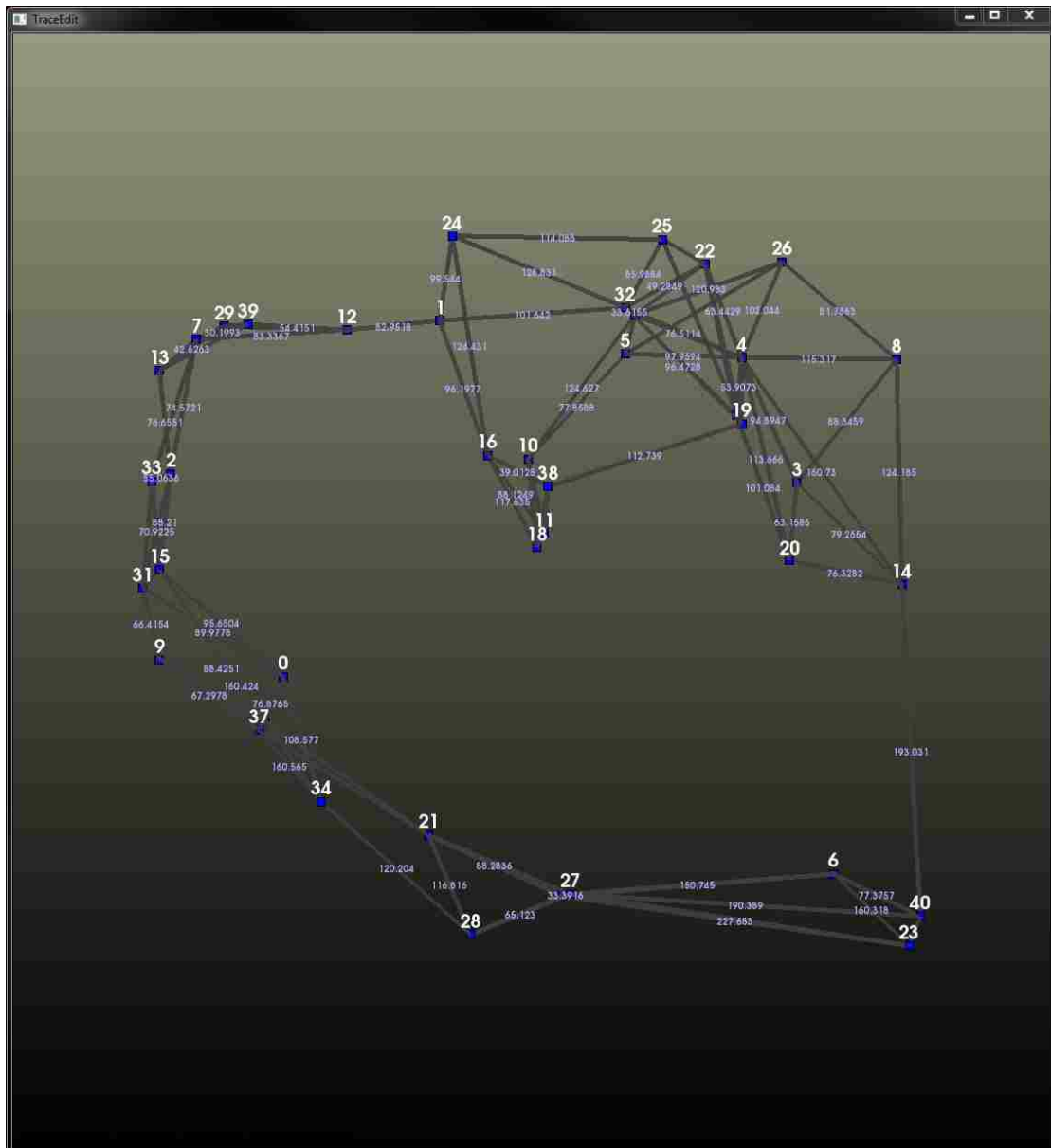
FIGURE 4.8: A cell to cell distance graph used for analyzing the data.

Analysis How to perform Coifman Harmonic Co-Clustering

1: Load the tracing result in the FARSIGHT Trace Editor

2: Perform the Trace Editors cellular analysis

Select Analysis → Cell Analysis

3: Perform a BiCluster Analyis

Select Analysis → BiClus Analysis

FIGURE 4.9: A more advanced analytical technique, this image displays a Coifman Plot showing harmonic co-clustering.

Chapter 5

Case Studies

This chapter looks at examples of the process described in Chapter 4 and demonstrates the robustness of this process to multiple conditions. The first section looks at analysis of control rats as a basis. The next example uses rats with optical coherence tomography (OCT) probes to demonstrate its effectiveness even with acute injury sites. The third example uses data from pigs exposed to a hypoxic environment during surgery and demonstrates its effectiveness for other species as well as broad disease-like states. The final example looks at an *in vitro* alginate study to demonstrate its use for mixed species and non-biological environments.

The last section of this chapter contains comparisons of the process (in rats) for the improved preparation techniques and the original methods as well as a statistical justification of their adoption.

As will be demonstrated, FARSIGHT is an excellent tool for quantifiable analysis of neuroimaging data. However, proper preparation of the data for use with FARSIGHT is essential. FARSIGHT is robust to the species and can be used for healthy tissue, acute injuries, and systemic issues as well.

5.1 Control Rats

5.1.1 Background and Significance

This section examines control rat data as basis for the other examples. The rats were prepared as described in Chapter 2, though no electrode was inserted into cortex, and the analysis performed as described in Chapter 4. This is the most basic walk-through of the process from start to finish.

5.1.2 Process and Results
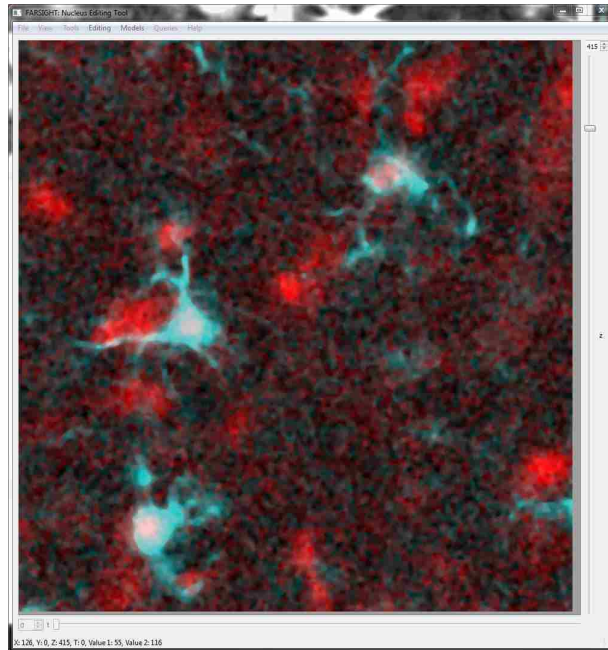
FIGURE 5.1: The basic image of control rat cortex used for analysis.
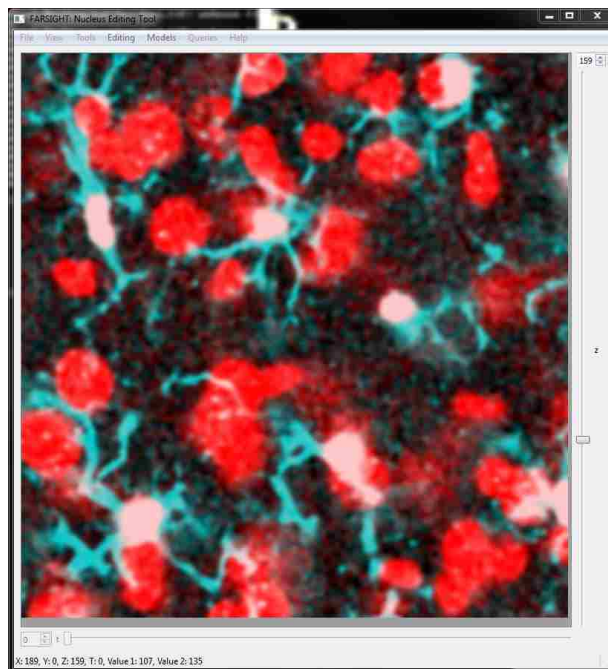
FIGURE 5.2: The results of the preparation phase on the control rat basic image. This image has a notably improved SNR.
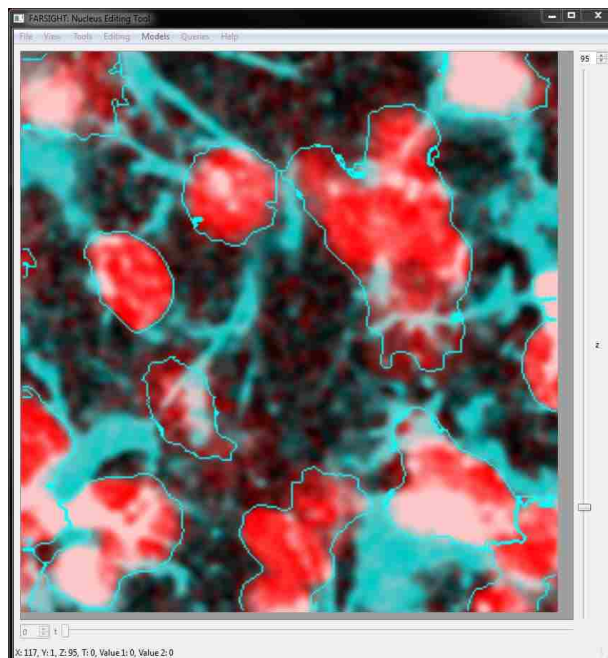
FIGURE 5.3: Several nuclei are clearly identified in this image after the segmentation algorithm is run in FARSIGHT.
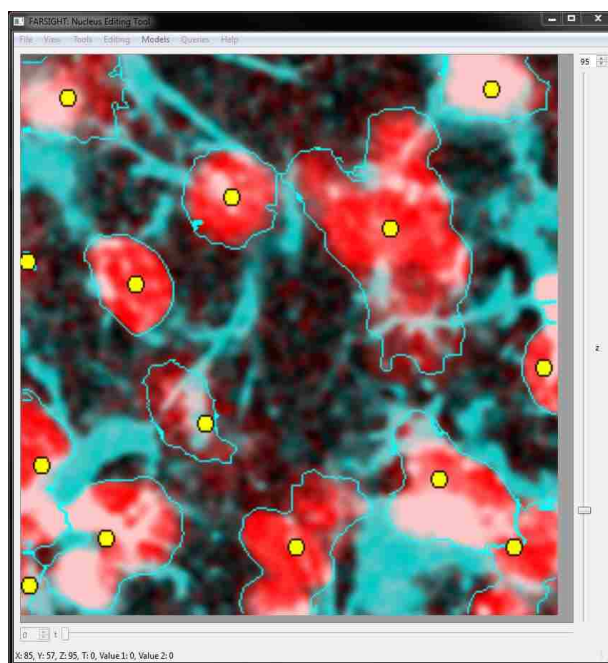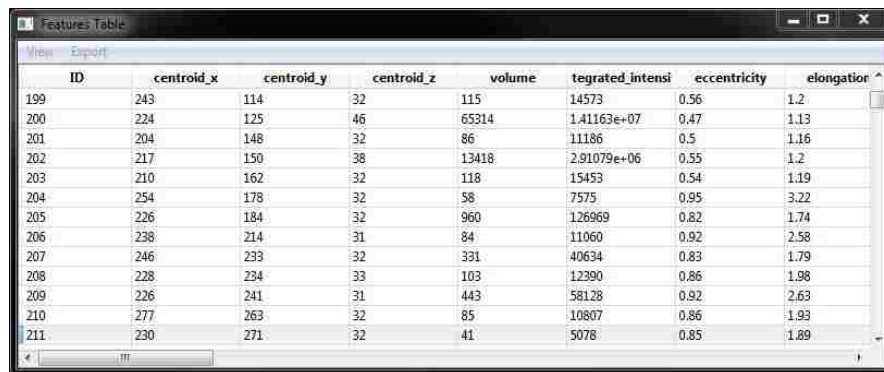
FIGURE 5.4: Object centroids are superimposed over the segmentation results, with colors corresponding to the classification of each cell. In this image, none of the nuclei belong to microglia.

FIGURE 5.5: The basic steps of analysis result in a feature table as shown. From this it is possible to determine cell counts for the total number of cells as well as each classification. Given that we known the area of this image field, we can also calculate the density of the cells given this information.

5.2 Rats with OCT Probes

5.2.1 Background and Significance

This section examines rat cortex after implantation with an optical coherence tomography probe and traditional neuroprosthetic electrode. This is a local injury to the cortex and shows that the methods still work in the presence of acute injury. The goal of this project was to use the OCT probe to observe and measure the tissue surrounding the electrode after implantation. Thus, the analytical process in this section is expanded from that of the control rats as localization and morphology of cells are relevant factors to the results of the experiment.

5.2.2 Process and Results
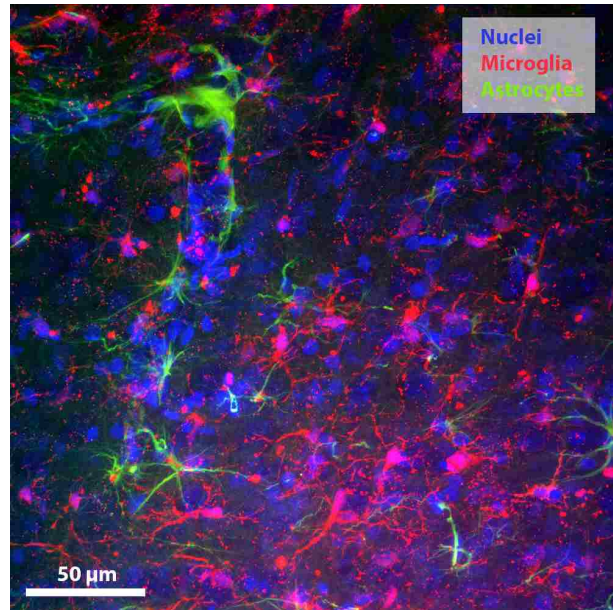
FIGURE 5.6: The starting image of rat cortex. This particular field does not contain the neuroprosthetic implant nor the OCT probe but is taken from a region between the two. This basic image still contains the astrocyte channel, which is removed for simplicity in the following images. It is important to note that this image is a maximum projection of the channels whereas images in FARSIGHT typically show a single z-slice.
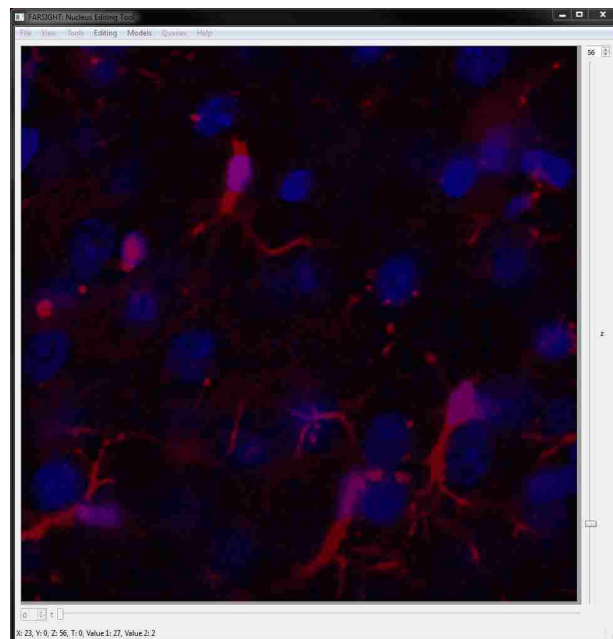
FIGURE 5.7: The image after filtering and background subtraction, shown in FAR-SIGHT this time. As well as showing only a single z-slice in FARSIGHT, the scale has been adjusted to adequately show the results of the process.
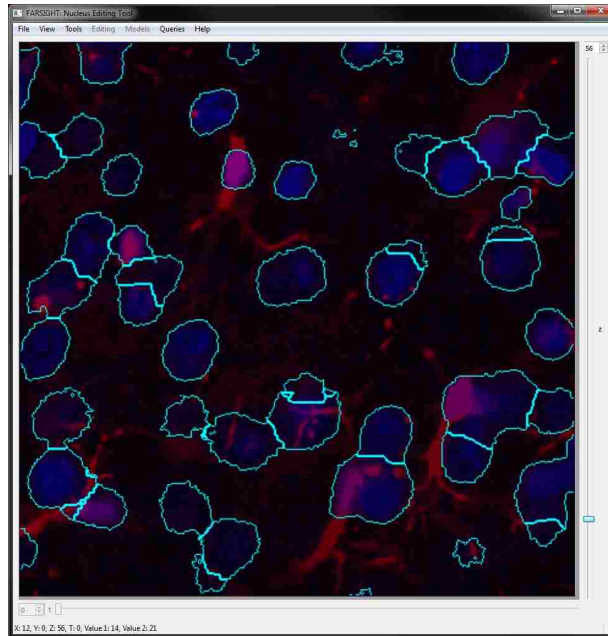
FIGURE 5.8: The results of the segmentation algorithm from FARSIGHT. The limitations of displaying a 3-dimensional environment in a 2-dimensional image are evident in this image, as nuclei that are distinct in the z-dimension appear to be overlapping in the x- and y-dimensions.
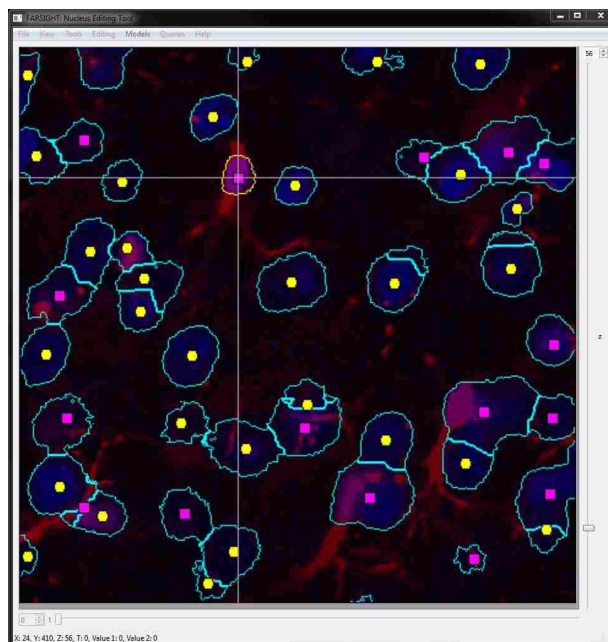
FIGURE 5.9: Results of classification within FARSIGHT. The color of the centroids superimposed on this image correspond to the classification of the cell. Magenta centroids indicate the nuclei belongs to a microglia while yellow indicates some other cell type. Note that the classes in this training session were "Microglia" and "Not Microglia" so we cannot assume that yellow centroids indicate neurons or any cell type specifically.
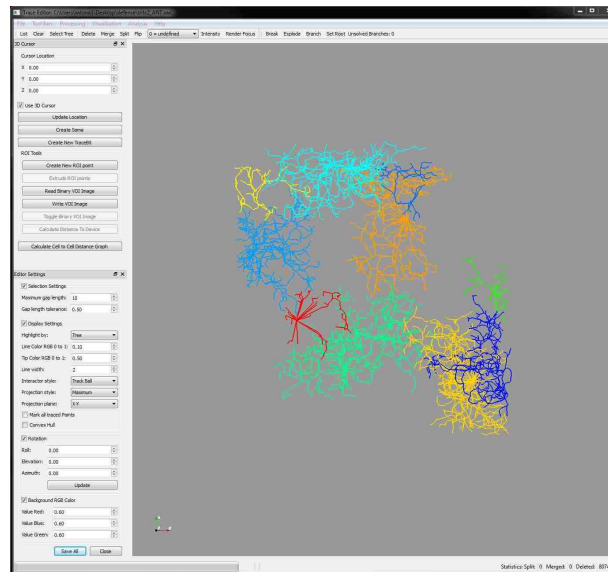
FIGURE 5.10: Tracing results shown in the FARSIGHT Trace Editor. Due to the extremely complex 3-dimensional nature of these glial arbors, only a portion of the cell tracings are shown. Further 3-dimensional manipulation of these tracings will show that several clusters of microglia are color-coded in the image.
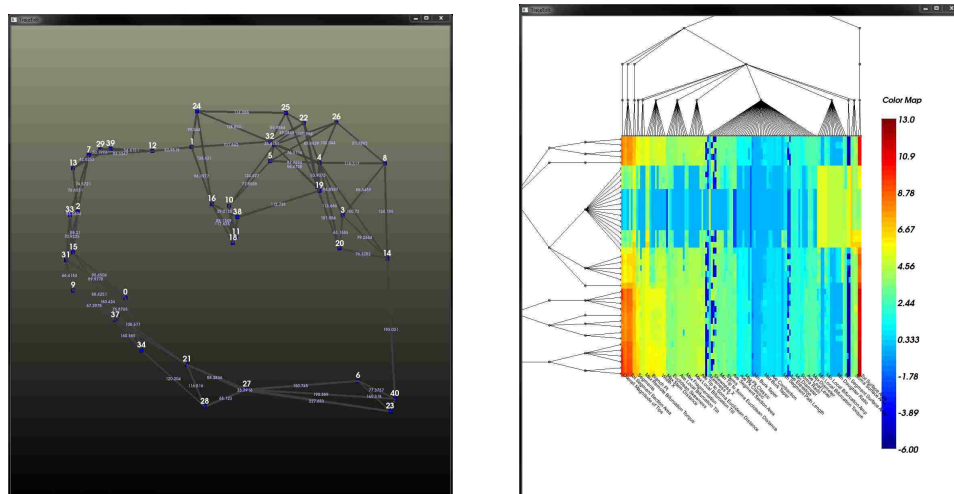
(a) Cell to Cell Distance Graph

(b) Coifman Plot

FIGURE 5.11: Further analysis of the tracing data can be performed in the FARSIGHT Trace Editor. These images show the results of these analyses, including a cell to cell distance graph and a Coifman plot. These tools can help identify glial activation states and map their spatial distributions [31].

5.3   Hypoxic Pigs

5.3.1   Background and Significance

This section examines pig cortex that was exposed to a hypoxic environment during a surgical intervention. Whereas the the OCT probe represents an acute injury, the hypoxic environment is a systemic insult with global effects. This sections demonstrates FARSIGHTs ability to analyze more than just local injury sites, and in fact it can be used for almost any kind of neuroscientific inquiry.

5.3.2   Process and Results
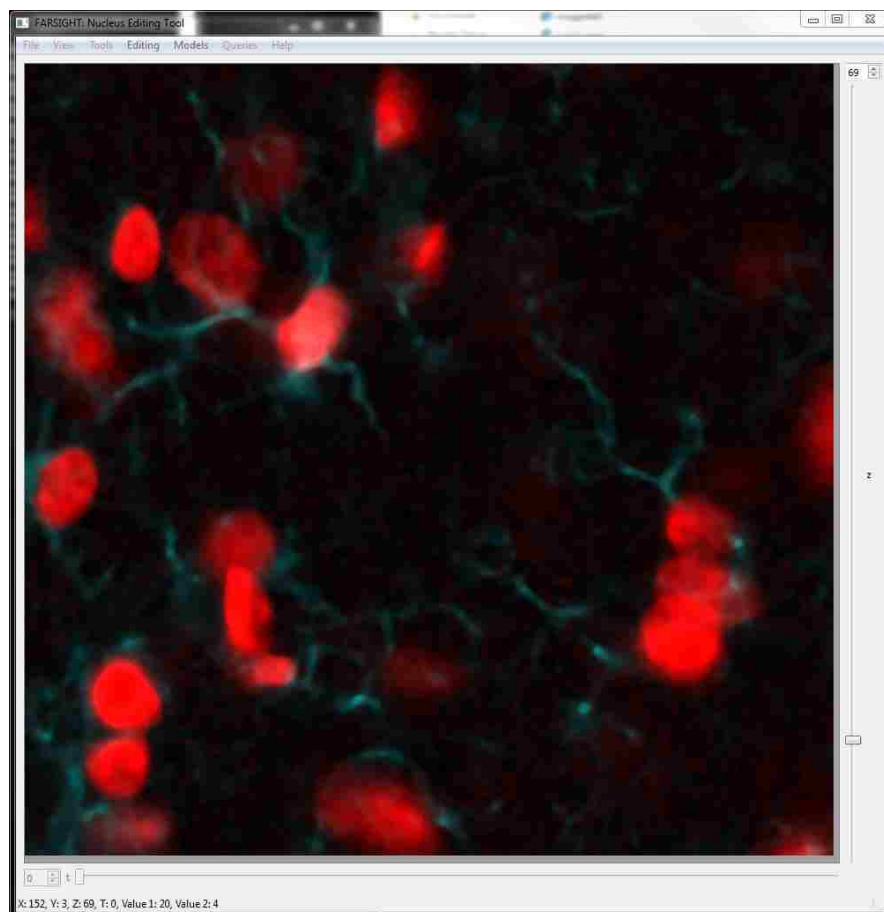
FIGURE 5.12: This is the basic image taken from the pig study, this time displayed after the preparation phase of filtering and background subtraction has been completed.
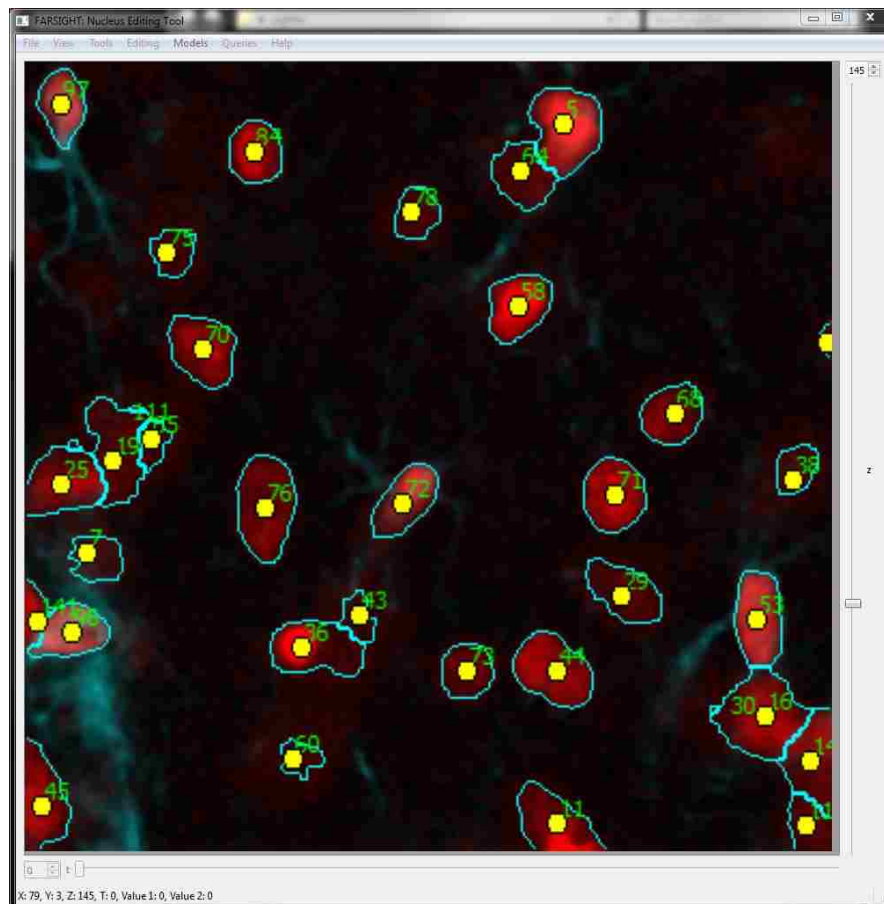
FIGURE 5.13: Results of the segmentation and classification algorithms in FARSIGHT. In this image, the object IDs are superimposed as well as the centroids.

ID	centroid_x	centroid_y	centroid_z	volume	tegrated_intensi	eccentricity	elongation
251	1	162	420	343	18076	0.76	1.54
252	72	313	220	333	18843	0.97	3.96
253	359	364	7	297	20669	0.94	3.01
254	363	366	210	281	16814	0.98	5.58
255	1	206	221	272	14736	0.71	1.42
256	356	364	28	267	18946	0.86	1.96
257	40	327	0	263	20043	1	13.08
258	367	365	29	205	13982	0.71	1.43
259	357	364	30	192	12713	0.94	2.84
260	348	93	0	173	10228	1	10.98
261	373	151	293	129	6618	0.67	1.34
262	216	338	0	118	6790	0.99	9.39
263	59	366	80	116	6551	0.92	2.48

FIGURE 5.14: The first step in analysis for this project was generating cell counts and calculating cell densities. This is accomplished through the feature table generated by FARSIGHT. In this field there are 263 cells.
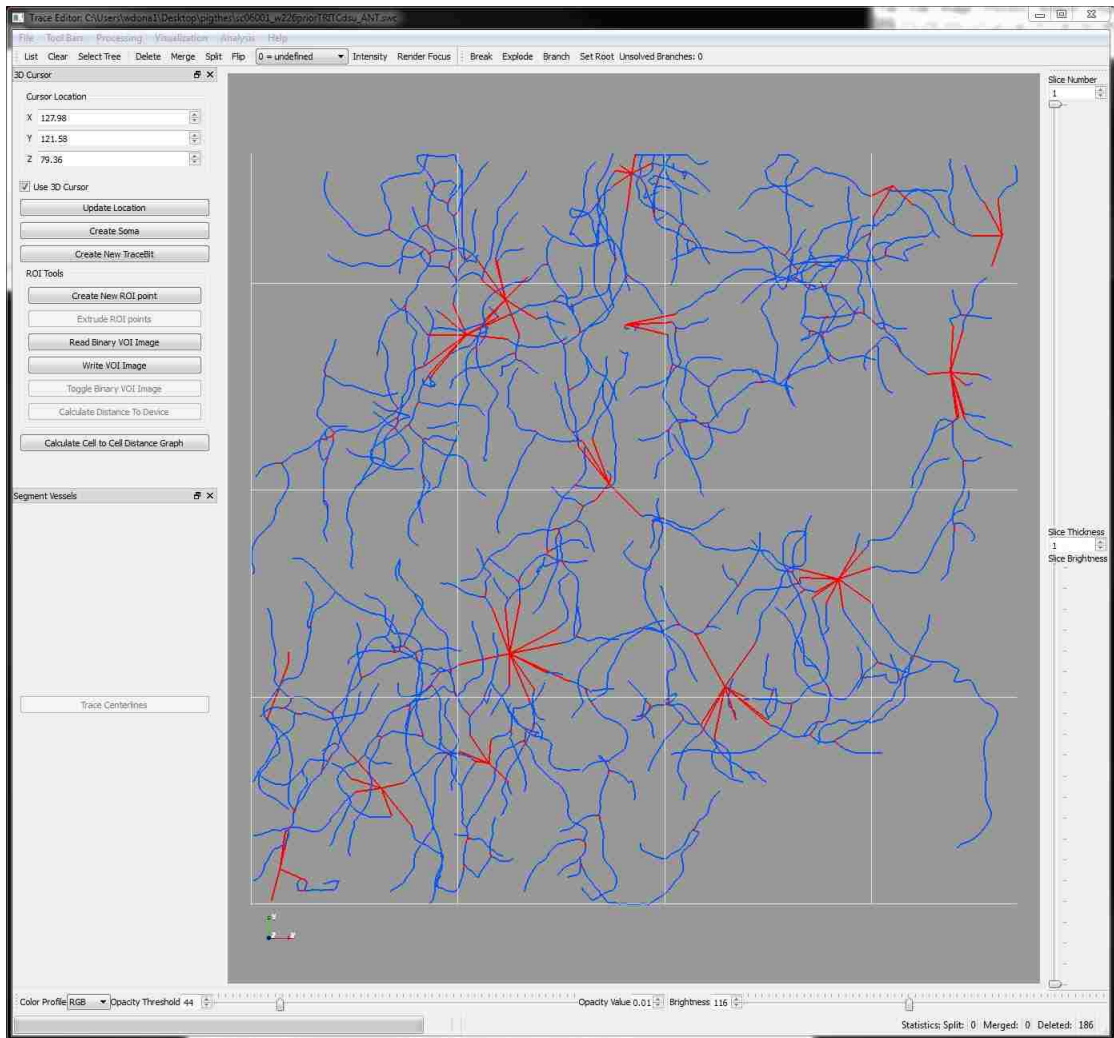
FIGURE 5.15: The results of tracing the pig data is shown here. This time, color-coding was forgone to illustrate the representation of individual cells more clearly. The red segments represent the soma of the cell, with their origin at the centroid. The blue segments represent the glial processes. Again, only some of the cells are still displayed in this image due to the difficulty of displaying the complex 3-dimensional arbors.

5.4   *In Vitro* Alginate Model

5.4.1   Background and Significance

This section examines an *in vitro* alginate model that was designed to be the same density and stiffness as human cortex. The alginate contains astrocytes harvested from rat pups and microglia harvested from a cell line at concentrations mimicking cortical levels. The goal of this experiment was to provide a strong *in vitro* model for hydrocephalus shunt studies. Here it is used to illustrate FARSIGHTs use for samples with mixed species of cells or taken from non-biological environments.

5.4.2 Process and Results
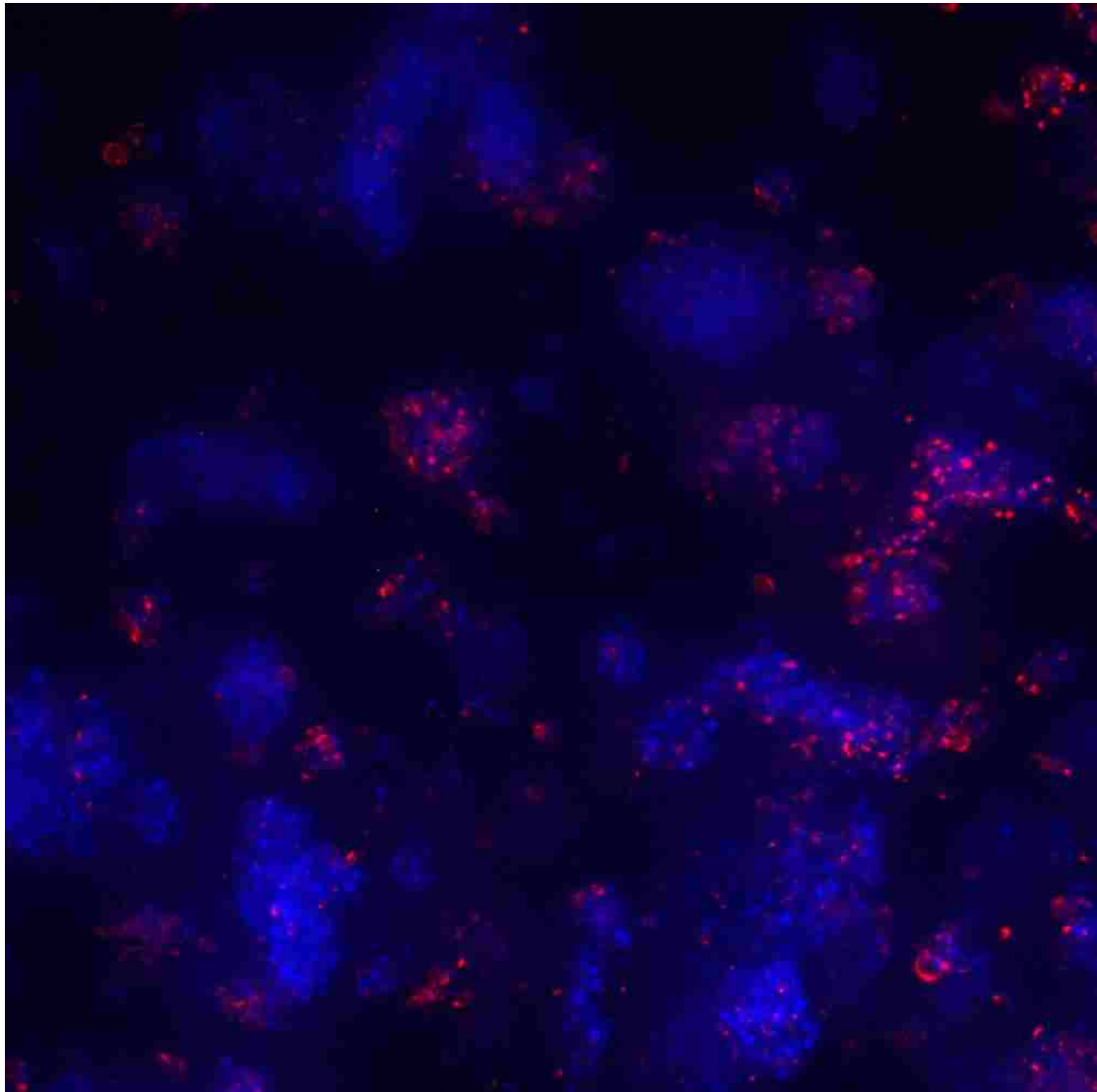
FIGURE 5.16: The basic image used for analysis in the alginate study. This is a maximum projection of the nuclear and microglia channels, rather than a single z-slice. Also of interest, while all the other images presented in this paper were taken at 30x, this project imaged at 10x. This, combined with the clustering evident in this image, does make resolution of individual nuclei harder in FARSIGHT and increases the runtime and reduces the certainty of the results. This means also that the preparation stage must be carefully examined to ensure that the cells can still be clearly identified.
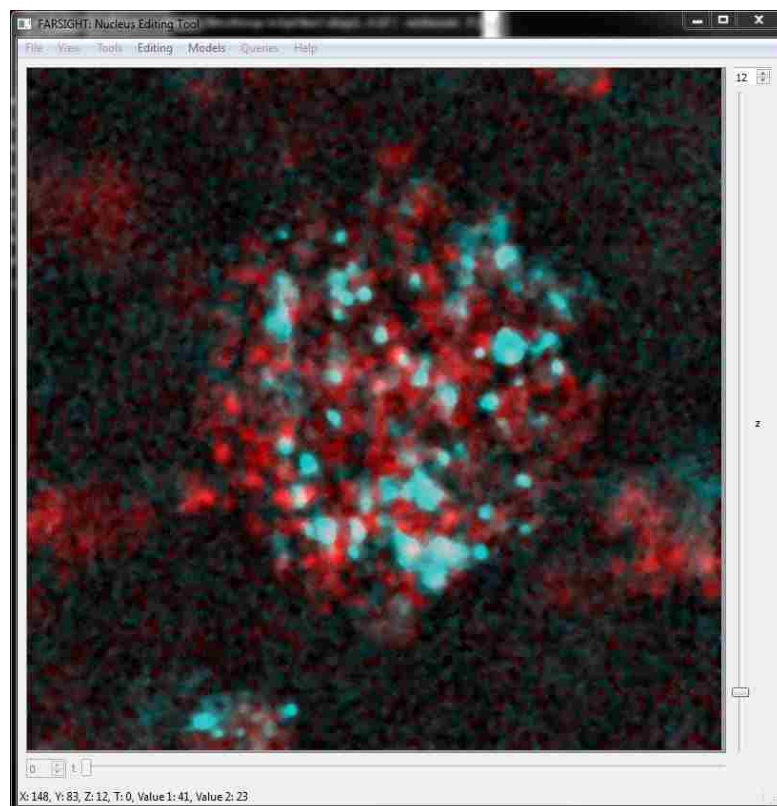
FIGURE 5.17: Alginate image loaded into FARSIGHT after preparation. The scale has been adjusted to better show the clustering of cells. There remains a greater amount of noise in this image, as the filtering results were adjusted to make sure that individual nuclei can still be resolved. Although they appear closely clustered and indistinct in this slice, the cells can still be resolved due to their separation in the z-dimension. Lastly, note that due to the abundance of cells in this image and their close spacing, displaying the results of segmentation and classification phase cannot be meaningfully displayed in a 2-dimensional format such as this.

FIGURE 5.18: The goal of these imaging studies was to verify cell counts and cell densities and determine if the alginate model was suitable for *in vitro* studies. To this extent, the cell count shown and the density calculations are the most important analytical measurements. Due to the limitations of 10x images, further analysis in FARSIGHT is not feasible at this time as there is not enough resolution for tracing.

5.5 Preparation Comparisons

In this final section I demonstrate the beneficial effects of the preparation techniques described in Chapters 3 and 4 with visual examples as well as a statistical analysis.
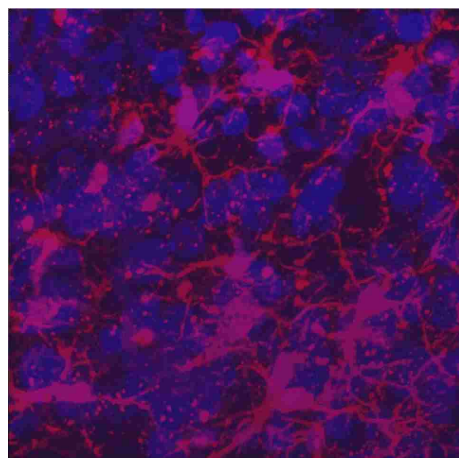
(a) Good Preparation
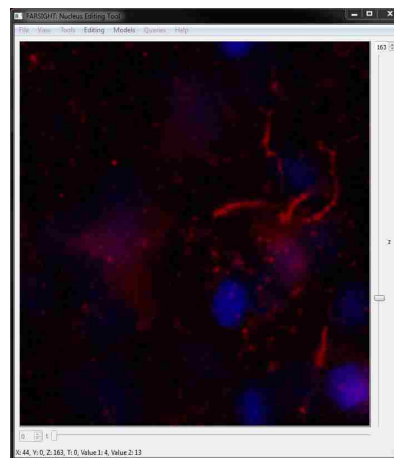
(b) Bad or No Preparation

FIGURE 5.19: Comparison of the preparation phase. Figure A shows a maximum projection of nuclear and microglia channels after undergoing proper filtering and background subtraction as described in this paper. Note there is very little high frequency noise and the background levels are low for a projection such as this. Figure B shows an image that was not filtered nor prepared for use with FARSIGHT. Significant high frequency noise is still evident in this picture as well as out-of-focus information contributing to an elevated background level even in a single slice as shown.
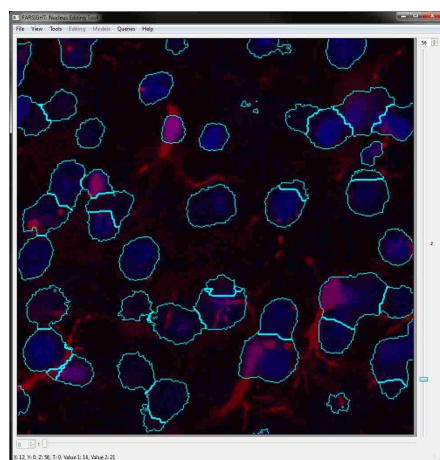
(a) Segmentation After Correct Prepartion
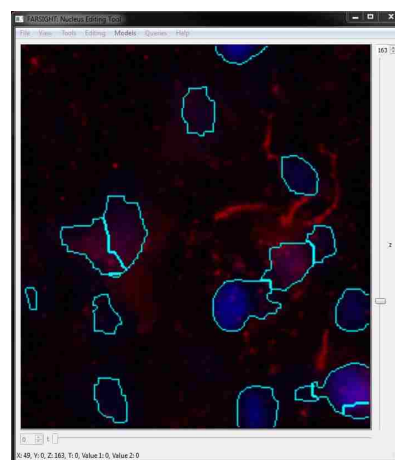
(b) Segmentation After Bad Prepartion

FIGURE 5.20: Comparison of the segmentation results. Figure A shows promising segmentation results after proper preparation. Figure B does not correctly segment the nuclei. The edges do not match the image, likely due to the noise levels. Additionally, the leftmost outline does not correspond with the location of a nucleus but rather with a level of higher background signal.
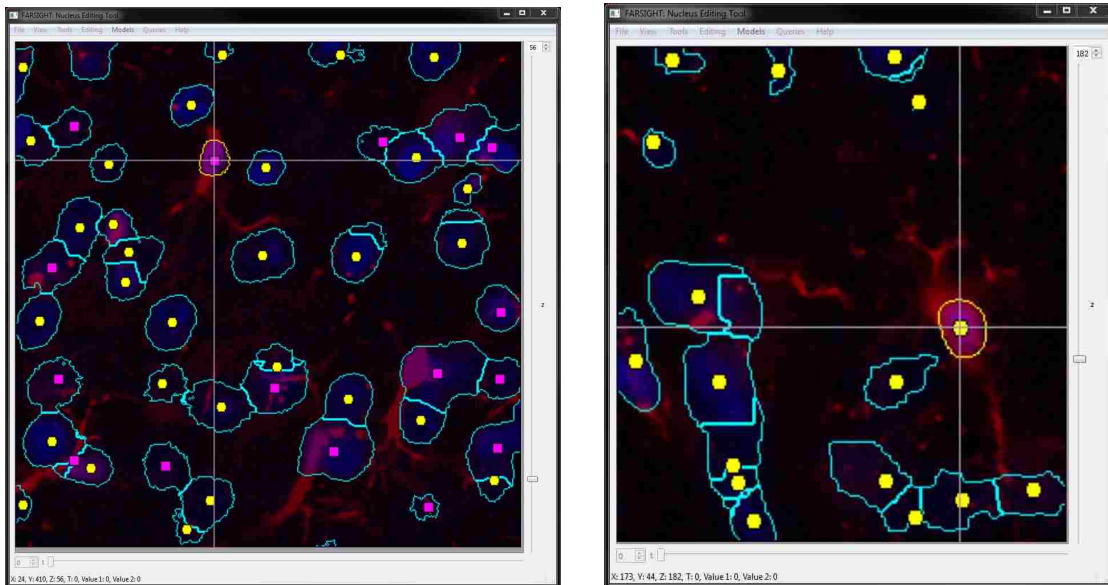
(a) Classification After Correct Preparation

(b) Classification After Bad Preparation

FIGURE 5.21: Comparison of classification results. Again, Figure A shows good results and accurate classification, while Figure B clearly misclassified the highlighted cell.
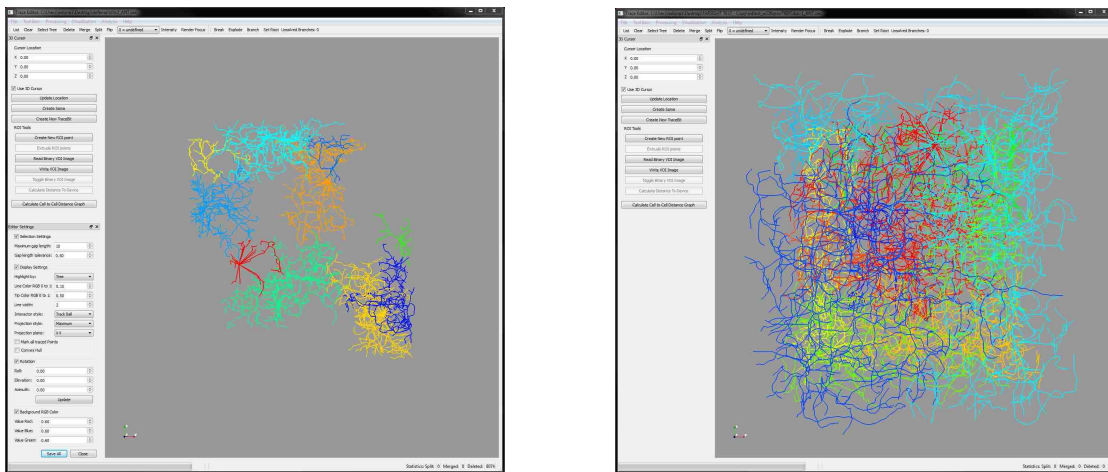
(a) Tracing Results After Correct Preparation

(b) Tracing Results After Bad Preparation

FIGURE 5.22: Comparison of tracing reults. Figure A shows good results as demonstrated before. Figure B on the other hand is completely unusable. The high noise levels contribute extremely negatively to the tracing results, often leading to wildly complicated arbors extending far beyond the actual cell. For comparison, the same number of cells are shown in each tracing. It is evident that at least some preparation work is necessary for high level analysis with FARSIGHT or the results become invalid.

5.5.1 Statistical Justification

This statistical test compares two different preparation methods (no preparation vs the method described in Chapters 3 and 4) to see how they affect the average time to complete the segmentation and tracing process. Each preparation method was applied to a large set of data consisting of some number of randomly selected images and then the segmentation and tracing process was be timed for each image in the data set, resulting in paired observations. From this I calculated an average time and standard deviation for each preparation method. I then performed a two-tailed paired t test (assuming equal variances) to determine how each method affected the total processing time. Note that it could be reasonable to perform either a Sign test or Wilcoxon rank sum test here if the only concern was which of the two algorithms were faster, but the actual time requirement can be compared to future improvements to these techniques. Thus a paired t test was used instead. Preliminary results show that Method #1 (no preparation) had a mean run time of 7600 seconds and Method #2 (as described) has a mean time of 7250 seconds. I determined that a sample size of 36 images in the data set will have 80% power to detect a difference of 60 seconds between the groups assuming a common standard deviation of 600 seconds based on a two-sided paired t test at a 0.05 significance level.

Based on these requirements, 36 images were generated and processed using each preparation method. The paired t test gave a p-value of 0.0214 for a mean decrease of 334 seconds using the method described in this paper. Thus there is a significant improvement in the time required for processing after use of the preparation techniques in this paper.

Chapter 6

Limitations and Future Work

6.1 Limitations of Current Implementation

The current implementation of these techniques, though much improved, is far from perfect. In this section I will discuss the limitations of the current methodology. I have identified the three greatest limitations to this approach: the necessary computational power, the time required for completion, and the size of the data. These three factors are all intimately linked together, so much so that they could easily be considered three facets of the same problem. However, they all deserve discussion in their own right, as solutions to ameliorate these limitations may address these factors independently.

6.1.1 Computational Power

FARSIGHT requires immense computational power, even by modern laboratory standards. For a single field captured for this document, the algorithm for segmentation alone required 72GB of memory (RAM) available. For larger data sets, especially montaged data sets, this requirement increases very rapidly and quickly outpaces the available computing resources of even well-equipped labs.

There are a number of possible avenues to pursue in order to reduce the computational power required. Here I will briefly discuss one of the immediate approaches that can be used to reduce the necessary resources. The memory required is dependent on the size of the image to be processed, so dicing the images into smaller segments will result in lower resource consumption. However this approach will have negative side effects. While dicing images results in smaller images and system loads, the same amount of data still must be processed so this approach will not result in reduced run-times. In fact, increasing the total number of images to be processed will likely increase the total time

due to overhead costs of loading images and the nonlinearity of algorithm run-times. This approach will also likely result in a less accurate analysis of the data. FARSIGHT uses an associative approach to its algorithms, especially for microglia tracing[2]. This means that if the image is cropped or diced in a manner such that a glial process is removed from the soma then the tracing algorithm will not be able to account for that missing data. So while dicing an image can be viable to reduce the system requirements to below the threshold of the computer at hand, it may not reduce the time requirements and can jeopardize the results of the analysis. I will discuss further methods for addressing this limitation in the Proposed Future Work section.

6.1.2 Time

Another major limitation is the time required to process the images. While the automation techniques in this paper reduce the man-hours involved, the process is still lengthy. Preprocessing an average data set requires up to 4 hours of computer run-time. The actual processing of that same data set will require up to 24 hours of run-time.

Reducing the time of processing does not have any quick fixes available unfortunately. This requires either significant monetary investment to upgrade computer hardware in order to handle larger image sizes or redevelopment of the algorithms behind FARSIGHT. As mentioned before, work on FARSIGHT is ongoing so improvements to algorithms will be incorporated into future releases and run-times will improve.

6.1.3 Size of Data

In order to achieve the best results possible, the data sets collected for these types of experiments often run up to 500GB in size. Not only is it necessary to store this much data for each experiment, that data must be easily accessible by the computer performing the analysis. Data transfer speeds thus become a limiting factor for significantly large data sets. Additionally, this affects the type of storage medium selected as read/write speeds should ideally not be the limiting factor in analysis.

Reducing the size of a data set is not a trivial task. Due to the complex nature of biological micro-environments, it is unwise to use lossy compression algorithms on the data. Ideally, all of the relevant data should be recorded and stored in a completely lossless format. This significantly limits the techniques we can use reduce the size of a data set. In fact, the only reasonable short term solution is simply to reduce the amount of data collected or processed. Focusing on hypothesis-driven testing can indicate the

regions of the data that are most reasonable to analyze. In this manner, it can be feasible to reduce the size of the data set.

6.2 Proposed Future Work

In order to combat these limitations I would propose several areas of future improvement. Development of FARSIGHT is an ongoing process and this project has illustrated the need for these improvements to be incorporated.

6.2.1 Further Automation

While the data preparation techniques discussed in this paper have been fully automated, the process within FARSIGHT requires further work to completely automate. Since the majority of this procedure can be completed through a command line interface, automation of the process is in the works.

6.2.2 Greater SNR Improvement

Signal-to-noise ratio compares the level of a desired signal to the level of background noise. As we discussed previously in this document it is a good measure for the success of the preprocessing techniques. While the preprocessing algorithm developed for this project greatly improves the SNR and results, it is likely that further study will result in yet greater improvements to the SNR and alterations to the algorithm.

6.2.3 Application to Astrocytes

As it stands, the tracing techniques discussed in this document apply only to microglia. Due to the nature of the algorithm these do not directly translate to astrocytes[2]. Further development is necessary to include these cell types. To ensure that FARSIGHT is a usable tool in the field of neuroscience it is necessary to include analytical techniques to address questions concerning all the major cell types.

6.2.4 Distributed Computing

Perhaps the most important development in addressing the limitations of this approach is implementing distributed computing methods. Distributed computing involves dividing

a problem into many smaller tasks which can each be solved by one or more computers in a network of computers which communicate by message passing[32]. To take advantage of distributed computing techniques will require modification of the algorithms to implement parallelized code. The amount of the process that can be parallelized will directly affect the decrease in run-time in accordance with Amdahl's law[33]. Additionally, the smaller divided tasks of a distributed approach will reduce the requisite computational power of any given node in the network. For the current performance-to-price ratios of hardware components, at this level of computation it is often more economically feasible to add another node into a computer network than to upgrade the hardware of a single workstation. Furthermore, implementation of parallelized code will allow researchers to outsource the computational portion of analysis to a third party distributed computing network rather than spend the capital to build their own computer infrastructure[34]. Thus, parallelization and distributed computing will ease the burden of computational limitations as well as potentially dramatically reduce the run-time of the process.

Chapter 7

Conclusion

The goal of this paper was to present a comprehensive method for analyzing neuroimaging data in an automated manner. Neuroimaging techniques are essential to both researchers and medical professionals, but analysis of these images is often subjective and exceptionally difficult. This provided the motivation for an automated technique that could be easily adopted by any researcher with a robust set of tools for generating quantifiable measurements for neuroscientific inquiries. The application of digital signal processing methods to prepare data in combination with the powerful FARSIGHT toolkit enable an accessible, efficient, and automated process described in this paper.

The second chapter in this paper sets the stage for the analytical process by discussing the experimental setup and the overall process that an experimenter might go through in order to complete this type of analysis. This is focused on the baseline example of a rat model implanted with a neuroprosthetic electrode, including the surgical procedure and immunohistochemical labeling.

The third chapter discussed the implications of signal-to-noise ratio as a measurement for image analysis and a marker for the success of the preparation stage. A mathematical approach to calculating the SNR was described and then the image analysis techniques, specifically median filtering and background subtraction, were introduced. This provided a fundamental justification for parts of the process described in the next chapter.

Chapter 4 detailed the exact process used in analysis of neuroimaging data. The preparation stage consisted of the digital signal processing techniques to improve the SNR and generation of the XML files necessary to interact with FARSIGHT. The processing stage itself was completed in FARSIGHT as well, beginning with segmentation, leading into training the classifier, and finally concluding with classification of cells and tracing

of glial cells for each image. The final stage contains the actual analysis of the quantification generated in the previous stages. Analysis depends explicitly on the experiment in question, but the techniques can be as simple as a cell count or as involved as a Coifman harmonic co-clustering plot.

Chapter 5 showed case studies of the methodology presented herein, demonstrating its effectiveness and robustness. The basic model used was control rat cortex, while rats with OCT probe implants demonstrated acute injuries, hypoxic pigs demonstrated global effects, and an alginate model showed the robustness of the model to multiple species and *in vitro* environments. This chapter also included comparisons of various stages in the process with proper data preparation techniques and without, as well as a statistical validation of the improved results.

Finally, chapter 6 detailed the limitations of this analytical method and proposed the next steps to be taken in the development and use of FARSIGHT. This method is still hampered by the computational requirements, size of the data sets, and run-time of the algorithms. Further work in automation as well as greater SNR improvements, more robust cellular applications, and especially implementation of distributed computing methods will help to address these limitations and contribute positively to the use and adoption of FARSIGHT by researchers everywhere. Automated comprehensive analytical techniques such as this are likely to play a major role in neuroscientific experiments in both the near and far future.

Appendix A

Image Definition Template

```
<Image>
    <file chname="nuc" r="0" g="0" b="255">path/to/file/directory/dapi.tif</file>
    <file chname="mig" r="255" g="0" b="0">path/to/file/directory/tritc.tif</file>
    <file chname="ast" r="0" g="255" b="0">path/to/file/directory/gfp.tif</file>
</Image>
```

Image Definition Template

Appendix B

Process Template

```
<ProjectDefinition name="Project Name">
    <Inputs>
        <channel number="0" name="nuc" type="NUCLEAR" />
        <channel number="1" name="mig" type="MICROGLIA" />
    </Inputs>
    <Pipeline>
        <step name="NUCLEAR_SEGMENTATION" />
        <step name="FEATURE_COMPUTATION" />
        <step name="RAW_ASSOCIATIONS" />
    </Pipeline>
    <NuclearSegmentationParameters>
        <parameter name="high_sensitivity" value="0.00" />
        <parameter name="adaptive_binarization" value="0.00" />
        <parameter name="LoG_size" value="40.00" />
        <parameter name="min_scale" value="10.00" />
        <parameter name="max_scale" value="20.00" />
        <parameter name="xy_clustering_res" value="4.00" />
        <parameter name="z_clustering_res" value="3.00" />
        <parameter name="finalize_segmentation" value="1.00" />
        <parameter name="sampling_ratio_XY_to_Z" value="3.00" />
        <parameter name="refinement_range" value="6.00" />
        <parameter name="min_object_size" value="100.00" />
    </NuclearSegmentationParameters>
    <AssociationRules>
        <AssociationRule Name="mig_TOTAL" SegmentationSource="NUCLEAR"
        Target_Image="mig" Outside_Distance="0" Inside_Distance="10"
        Use_Whole_Object="True" Use_Background_Subtraction="True"
        Use_MultiLevel_Thresholding="False" Number_Of_Thresholds="1"
        Number_Included_In_Foreground="1" Association_Type="TOTAL" />
        <AssociationRule Name="mig_AVERAGE" SegmentationSource="NUCLEAR"
        Target_Image="mig" Outside_Distance="0" Inside_Distance="10"
        Use_Whole_Object="True" Use_Background_Subtraction="True"
        Use_MultiLevel_Thresholding="False" Number_Of_Thresholds="1"
        Number_Included_In_Foreground="1" Association_Type="AVERAGE" />
    </AssociationRules>
</ProjectDefinition>
```

Segmentation Process Definition Template

Appendix C

Process Template with Classification

```
<ProjectDefinition name="Project Name">
    <Inputs>
        <channel number="0" name="nuc" type="NUCLEAR" />
        <channel number="1" name="mig" type="MICROGLIA" />
    </Inputs>
    <Pipeline>
        <step name="NUCLEAR_SEGMENTATION" />
        <step name="FEATURE_COMPUTATION" />
        <step name="RAW_ASSOCIATIONS" />
        <step name="CLASSIFY_MCLR" />
        <step name="CLASS_EXTRACTION" />
    </Pipeline>
    <NuclearSegmentationParameters>
        <parameter name="high_sensitivity" value="0.00" />
        <parameter name="adaptive_binarization" value="0.00" />
        <parameter name="LoG_size" value="40.00" />
        <parameter name="min_scale" value="10.00" />
        <parameter name="max_scale" value="20.00" />
        <parameter name="xy_clustering_res" value="4.00" />
        <parameter name="z_clustering_res" value="3.00" />
        <parameter name="finalize_segmentation" value="1.00" />
        <parameter name="sampling_ratio_XY_to_Z" value="3.00" />
        <parameter name="Use_Distance_Map" value="1.00" />
        <parameter name="refinement_range" value="6.00" />
        <parameter name="min_object_size" value="100.00" />
    </NuclearSegmentationParameters>
```

Classification Process Definition Template

```
    <AssociationRules>
        <AssociationRule Name="mig_TOTAL" SegmentationSource="NUCLEAR"
        Target_Image="mig" Outside_Distance="0" Inside_Distance="10"
        Use_Whole_Object="True" Use_Background_Subtraction="True"
        Use_MultiLevel_Thresholding="False" Number_Of_Thresholds="1"
        Number_Included_In_Foreground="1" Association_Type="TOTAL" />
        <AssociationRule Name="mig_AVERAGE" SegmentationSource="NUCLEAR"
        Target_Image="mig" Outside_Distance="0" Inside_Distance="10"
        Use_Whole_Object="True" Use_Background_Subtraction="True"
        Use_MultiLevel_Thresholding="False" Number_Of_Thresholds="1"
        Number_Included_In_Foreground="1" Association_Type="AVERAGE" />
    </AssociationRules>
    <Classification_MCLR_Rules>
        <Classification_MCLR_Rule ClassColName="mg" ConfThreshold="0.50"
        TrainingFileName="path/to/file/mg_classifier.txt" />
    </Classification_MCLR_Rules>
    <ClassExtractionRules>
        <ClassExtractionRule Class_Name="mig" Class="2" />
    </ClassExtractionRules>
</ProjectDefinition>
```

Segmentation Process Definition Template - *continued*

Bibliography

[1] B Pakkenberg and HJG Gundersen. Total number of neurons and glial cells in human brain nuclei estimated by the disector and the fractionator. *Journal of microscopy*, 150(1):1–20, 1988.

[2] B Roysam, W Shain, E Robey, Y Chen, A Narayanaswamy, CL Tsai, Y Al-Kofahi, C Bjornsson, Ena Ladi, and Paul Herzmark. The farsight project: associative 4d/5d image analysis methods for quantifying complex and dynamic biological microenvironments. *Microscopy and Microanalysis*, 14(S2):60–61, 2008.

[3] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012.

[4] Ennio Pannese. The golgi stain: invention, diffusion and impact on neurosciences. *Journal of the history of the neurosciences*, 8(2):132–140, 1999.

[5] Angelo Mosso. Sulla circolazione del sangue nel cervello dell'uomo. 1882.

[6] Stefano Sandrone, Marco Bacigaluppi, Marco R Galloni, Stefano F Cappa, Andrea Moro, Marco Catani, Massimo Filippi, Martin M Monti, Daniela Perani, and Gianvito Martino. Weighing brain activity with the balance: Angelo mossos original manuscripts come to light. *Brain*, 137(2):621–633, 2014.

[7] Walter E Dandy. Ventriculography following the injection of air into the cerebral ventricles. *Annals of surgery*, 68(1):5, 1918.

[8] Anne G Osborn. *Introduction to cerebral angiography*. HarperCollins Publishers, 1980.

[9] Jiang Hsieh. Computed tomography: principles, design, artifacts, and recent advances. SPIE Bellingham, WA, 2009.

[10] Robert L Savoy. History and future directions of human brain mapping and functional neuroimaging. *Acta psychologica*, 107(1):9–42, 2001.

[11] Paul T. Callaghan. *Principles of nuclear magnetic resonance microscopy*. Oxford University Press, 1993.

[12] Scott A Huettel, Allen W Song, and Gregory McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[13] Cedars-Sinai. Ct brain perfusion, aug 2015. URL `http://www.cedars-sinai.edu/Patients/Programs-and-Services/Imaging-Center/For-Physicians/Neuroradiology/Images/CT-Brain-Perfusion2-11791.jpg`. [Online; accessed August 2015].

[14] ExtremeTech. Mri brain scan, aug 2015. URL `http://www.cedars-sinai.edu/Patients/Programs-and-Services/Imaging-Center/For-Physicians/Neuroradiology/Images/CT-Brain-Perfusion2-11791.jpg`. [Online; accessed August 2015].

[15] American Academy of Neurology. Pet brain scan, aug 2015. URL `http://www.neurology.org/content/69/9/871/F1.large.jpg`. [Online; accessed August 2015].

[16] Antonio Pinto and Luca Brunese. Spectrum of diagnostic errors in radiology. *World journal of radiology*, 2(10):377, 2010.

[17] HJG Gundersen and EB Jensen. The efficiency of systematic sampling in stereology and its prediction*. *Journal of microscopy*, 147(3):229–263, 1987.

[18] Robert J Schalkoff. *Digital image processing and computer vision*, volume 286. Wiley New York, 1989.

[19] Louis L Scharf. *Statistical signal processing*, volume 98. Addison-Wesley Reading, MA, 1991.

[20] John M Boone, Thomas R Nelson, Karen K Lindfors, and J Anthony Seibert. Dedicated breast ct: Radiation dose and image quality evaluation 1. *Radiology*, 221 (3):657–667, 2001.

[21] Perry Lawrence Johnson. Iso 9000: Meeting the new international standards. *McGraw-Hill, Inc.(USA), 1993,*, page 246, 1993.

[22] Steven M Kay. *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, volume 3. Pearson Education, 2013.

[23] Olivier Rioul and Martin Vetterli. Wavelets and signal processing. *IEEE signal processing magazine*, 8(LCAV-ARTICLE-1991-005):14–38, 1991.

[24] DRK Brownrigg. The weighted median filter. *Communications of the ACM*, 27(8): 807–818, 1984.

[25] University of Southampton. Median filter, aug 2015. URL `http://users.ecs.soton.ac.uk/msn/book/new_demo/median/Picture1.png`. [Online; accessed August 2015].

[26] Sean M Hartig. Basic image analysis and manipulation in imagej. *Current protocols in molecular biology*, pages 14–15, 2013.

[27] Max Gassmann, Beat Grenacher, Bianca Rohde, and Johannes Vogel. Quantifying western blots: pitfalls of densitometry. *Electrophoresis*, 30(11):1845–1855, 2009.

[28] ImageJ. Rolling-ball background subtraction, aug 2015. [Online; accessed August 2015].

[29] DH Szarowski, MD Andersen, S Retterer, AJ Spence, M Isaacson, HG Craighead, JN Turner, and W Shain. Brain responses to micro-machined silicon devices. *Brain research*, 983(1):23–35, 2003.

[30] Justin C Williams, Robert L Rennaker, and Daryl R Kipke. Long-term neural recording characteristics of wire microelectrode arrays implanted in cerebral cortex. *Brain Research Protocols*, 4(3):303–313, 1999.

[31] Murad Megjhani, Nicolas Rey-Villamizar, Amine Merouane, Yanbin Lu, Amit Mukherjee, Kristen Trett, Peter Chong, Carolyn Harris, William Shain, and Badrinath Roysam. Population-scale three-dimensional reconstruction and quantitative profiling of microglia arbors. *Bioinformatics*, page btv109, 2015.

[32] David Peleg. Distributed computing. *SIAM Monographs on discrete mathematics and applications*, 5, 2000.

[33] Mark D Hill and Michael R Marty. Amdahl's law in the multicore era. *Computer*, (7):33–38, 2008.

[34] Rajesh Raman, Miron Livny, and Marvin Solomon. Matchmaking: Distributed resource management for high throughput computing. In *High Performance Distributed Computing, 1998. Proceedings. The Seventh International Symposium on*, pages 140–146. IEEE, 1998.