EM image analysis for
neuron segmentation, thickness estimation, and synaptic bouton quantification

A thesis submitted to attain the degree of
Doctor of Sciences ETH Zürich
(Dr. sc. ETH Zürich)

presented by
Thanuja Dakshitha Ambegoda Liyana Arachchige
Master of Science in Computational Biology and Bioinformatics
Department of Computer Science, ETH Zürich

Date of birth
December 6th, 1984

citizen of
Sri Lanka

accepted on the recommendation of

Prof. Dr. Richard Hahnloser (ETH Zürich), examiner
Dr. Matthew Cook (ETH Zürich), co-examiner
Prof. Dr. Orcun Goksel (ETH Zürich), co-examiner

2017

Abstract

Electron microscopy (EM) enables a detailed view of neuroanatomical structures in the brain at nanometer resolution. Information uncovered by EM has been instrumental in providing valuable insights of neuroanatomy and its relationship to higher level functions such as learning and memory. Petabytes of EM image data have been acquired at research facilities around the world for neuroscience research. These datasets are being analyzed mainly to investigate neuronal connectivity patterns at circuit level and to observe changes in neuronal structures such as synapses, as part of biological hypothesis testing. Since manual annotation of these images by humans is an inefficient and tedious task, automatic and semiautomatic image analysis methods are being developed to make this workflow more efficient.

The first part of this dissertation provides methods for EM image analysis that can be applied at two different stages of a typical EM image processing workflow. Chapter 2 presents a method for indirect automatic estimation of section thickness in serial section EM (ssEM) using image statistics alone. This method can be applied before annotation of images as a post-acquisition step to have a better estimate of the image volume at the level of individual sections. Accurate quantification of section thickness is useful since the thickness of sections can deviate from the expected thickness as a result of imperfections in the cutting process. The method proposed allows more accurate 3D representations of structures obtained by ssEM and provides more accurate volumetric statistics. As part of this work, we have also produced the first publicly available dataset for validating section thickness methods by directly measuring section thickness using atomic force microscopy.

Chapter 3 describes a method to automatically segment neuron slices in individual (2D) EM sections. This method focuses on lowering topological errors that take place in automatic segmentation methods. Such errors are typically caused by low signal-to-noise ratio in the input images. Main topological segmentation errors are either merge errors where two adjoining neuron slices in the data are given the same label or split errors where one neuron slice is segmented as two different neuron slices. The proposed method lowers such errors by enforcing topological constraints for the segmentation task, which is formulated as an integer linear program. All constraints used are local constraints that give rise to global behaviors - for instance, enforcing loop closure for neuron membranes in the segmentation output. Such constraints along with an objective function incorporating prior probabilities of segmentation labels result in lower topological errors as shown in the results.

An image-driven analysis of neuroanatomical structure in the zebra finch brain is presented in part 2 of this dissertation. Male juvenile zebra finches are known to learn

a stereotyped song from a tutor bird. Chapter 5 presents an investigation of changes in the number of multisynaptic boutons (MSB) in brain area HVC of juvenile birds as a result of tutor exposure. Results obtained show that although there is a slight trend towards a transient increase of the number of MSBs (per cubic micrometer) just after tutoring started, this interference is not statistically significant. As a secondary observation, it was seen that MSBs which have symmetric synapses significantly outnumber MSBs having asymmetric synapses in all of the volumes analysed, irrespective of the experimental conditions.

The need for automatic and semi-automatic methods for image analysis is on the rise with the increase of images acquired for research. In general, the accuracy of automated methods that are used for EM image analysis still have a plenty of room for improvement. Methods provided in this dissertation have contributed towards improving both post-acquisition image processing and automated segmentation, both of which are important stages of the EM image analysis workflow in neuroscience.

Sommario

La microscopia elettronica (EM, dall'inglese electron microscopy) consente una visione dettagliata delle strutture neuroanatomiche presenti nel cervello, con risoluzione nanometrica. Si tratta di una tecnica che ha avuto un ruolo fondamentale nel fornire informazioni preziose sulla neuroanatomia e sulla sua relazione con funzioni ad alto livello come l'apprendimento e la memoria. Presso le strutture di ricerca in tutto il mondo, sono stati acquisiti petabyte di dati estrapolati da immagini di microscopia elettronica per effettuare ricerche nel settore delle neuroscienze. Nel quadro della verifica dell'ipotesi biologica, questi dati sono esaminati principalmente per studiare i pattern di connettività neuronale a livello di circuito e per eseguire un'analisi quantitativa delle variazioni in strutture neuronali come le sinapsi. In considerazione del fatto che l'annotazione manuale eseguita da esperti umani è un'attività inefficiente e tediosa, sono in via di sviluppo metodi automatici e semiautomatici di analisi delle immagini per rendere più efficiente questo processo.

La prima parte di questa tesi fornisce metodi di analisi delle immagini di EM che possono essere applicati a due fasi differenti del tradizionale flusso di lavoro. Nel capitolo 2 viene presentato un metodo per la stima automatica indiretta dell'anisotropia, insieme al piano di imaging e allo spessore delle sezioni nella EM con sezione seriale (ssEM). Tale metodo può essere applicato prima dell'annotazione delle immagini; si tratta di una fase post-acquisizione che consente di ottenere una stima più precisa del volume delle immagini a livello di ciascuna sezione, ed è utile in quanto lo spessore delle sezioni può essere differente rispetto allo spessore previsto a causa di imperfezioni nel processo di taglio. Il metodo proposto consente di avere a disposizione rappresentazioni 3D più accurate delle strutture ottenute tramite la ssEM e stime più precise dei volumi. Come parte di questo lavoro, abbiamo prodotto il primo set di dati accessibile al pubblico dei metodi di validazione dello spessore della sezione, misurando direttamente lo spessore della sezione tramite la microscopia a forza atomica.

Il capitolo 3 descrive un metodo per segmentare automaticamente le sezionidi neuroni nelle singole (2D) sezioni della EM. Questo metodo ha come obiettivo la riduzione degli errori topologici che si verificano nelle procedure di segmentazione automatica, causati generalmente da un rapporto segnale-rumore (SNR, dall'inglese Signal-to-noise ratio) basso nelle immagini di input. I principali errori topologici di segmentazione possono essere errori di unione, nel caso in cui sia assegnata la stessa etichetta a due sezioni di neuroni adiacenti nei dati, o errori di separazione, nel caso in cui una sezione di neurone sia segmentata in due diverse sezioni di neurone. Il metodo proposto consente di ridurre tali errori applicando dei vincoli topologici all'attività di segmentazione, che è formulata come un programma lineare intero. Tutti i vincoli utilizzati sono vincoli locali che danno

origine a comportamenti globali portando, ad esempio, a una chiusura dell'anello (loop closure) per le membrane dei neuroni che circondano le sezioni di neuroni nell'output di segmentazione.

Nella seconda parte di questa tesi, viene presentata un'analisi basata sulle immagini relativa alle variazioni nella struttura neuroanatomica del diamante mandarino. È noto che i diamanti mandarini giovani di sesso maschile apprendono un canto stereotipato da un uccello maschio tutore, che in un ambiente naturale è il loro padre. Nel capitolo 5 viene presentato uno studio delle variazioni nel numero di bottoni multisinaptici (MSB, dall'inglese multisynaptic boutons) in un'area del cervello detta HVC (usata come nome proprio). I risultati ottenuti mostrano che, nonostante ci sia una leggera tendenza a un aumento transitorio del numero di MSB (per micrometro cubo) subito dopo l'inizio della fase di tutoring, tale aumento non è significativo. Come osservazione secondaria, si è notato che i MSB che presentano sinapsi asimmetriche sono estremamente più numerosi dei MSB con sinapsi simmetriche in tutti i volumi esaminati, a prescindere dalle condizioni sperimentali.

L'esigenza di introdurre nuovi metodi automatici e semiautomatici per l'analisi delle immagini è sempre più forte, a causa dell'aumento delle immagini acquisite a scopo di ricerca. In generale, ci sono ancora molti margini di miglioramento per quanto concerne la precisione dei metodi automatici utilizzati per l'analisi delle immagini di EM. I metodi proposti in questa tesi hanno contribuito a un miglioramento dell'elaborazione delle immagini post-acquisizione, cos come della segmentazione automatica: entrambe rappresentano fasi importanti del flusso di lavoro dell'analisi delle immagini di EM nel settore delle neuroscienze.

Acknowledgement

for organizing the annual events and the efforts to keep all PhD students in the program on track.

Stressful times are an inherent part of doctoral studies possibly because it is supposed to help expand one's limits. Especially now that I'm done with it, I think it might have been awkward to finish a PhD without much stress. Irrespective of mounting work pressure at times, I was able to maintain a relaxed and largely positive attitude towards not only work, but also life in general thanks to my colleagues who are also good friends. A short and obviously non-exhaustive list of my good friends at INI who have always been supportive would be: Jakob Buhmann, Gaby Michel, Julien Martel, Sergey Burnos, Jonathan Binas, Turlough Neary, Julia Buhmann, Federico Corradi, Lorenz Mueller and Richard George. I was lucky to have good friends outside the institute as well who have been equally supportive. Justin, Rounak and Julian are some of them. Thank you for making my stay in Zurich highly enjoyable. That could mainly be due to an interesting sense of humor that all of us share, in addition to their kindness.

Thanks to my aunt and uncle Bruno for hosting me at their beautiful house in the alps several times over the last few years. I haven't been able to visit as frequently as I would have liked to, but those visits have always been a breath of fresh air.

Many thanks to my flatmates in Zurich, Gian-Alberto and Slavtcho, for helping to maintain a peaceful and relaxed atmosphere at our shared apartment. I would also like to thank my cofounders Julian and Bo, for their patience and understanding in the final phase of my thesis work, especially because during the last few months I have been distributing my time between our new startup and wrapping up this dissertation. Also thanks to Federica Assenza and Federico Corradi for proofreading the Italian version of the abstract, which albeit in a small way, added a bit of an authentic Italian flavor, which I have been enjoying through out my stay in Europe.

Special thanks go to my two brothers who are always just a phone call or a text message away, although we are thousands of miles apart. Thank you for keeping our parents happy all the time and being there for me always. Last but not least, my heartfelt gratitude goes out to my parents who are my ultimate support system and primary source of inspiration. I was able to reach where I am today only because you let me stand up on your shoulders.

Contents

Chapter 1

Introduction

Understanding how the brain works requires analysing its structure and behavior at different conditions and spatial scales. Several decades ago neuroscientists began to analyze the internal structure of the brain at a macroscopic level using histological methods that lead to identification of brain regions. Work of Cajal and Golgi in analyzing the brain at a microscopic scale around early 1900s formed the foundations of modern neuroscience. It was Cajal who first proposed based on his observations that the nervous system is made up of individual cells which were later named neurons. This lead to finding connected networks of branching neurons in the brain.

Modern imaging techniques have enabled neuroscientists to observe anatomical properties both in static and dynamic settings at different spatial resolutions. For instance, fMRI imaging [59] allows spatio-temporal changes in 'blood oxygen level dependant contrast' to be observed, which is known to be correlated with neuronal activity. This is a non-invasive technique applicable to human subjects giving rise to images with a spatial resolutions around a cubic millimeter.

Developments in optical microscopy have enabled imaging at a resolution in the order of hundreds of nanometers. Combined with fluorescent imaging techniques, light microscopy is now capable of directly resolving neuron connectivity with the use of neurotracers [78]. Neurotracers result in fluorescence signals in neurons that have synaptic connections between them. Two-photon microscopy [15] is an instance of optical microscopy combined with fluorescent microscopy that enables imaging of living tissue up to about 1mm in depth.

Recent developments in 'expansion microscopy (ExM)' [11] enables conventional light microscopy to observe structures much smaller than what is allowed by the limitations of wavelength of light. The sample preparation protocol in ExM increases the physical size of the specimen up to five times its original size using a polymer. Therefore the traditionally cited resolution of 200 nm can be brought down to 40 nm which suffices to visualize synaptic vesicles.

Electron microscopy (EM) is an invasive technique that cannot be used on living cells as it requires biological tissue to be prepared according to special protocols that include staining the specimen with molecules that contain heavy metals. Currently, this is the only imaging technique that provides resolutions high enough to visualize neuroanatomical structures such as synapses having a spatial magnitude spanning a only few nanometers.

Neurons are able to transmit and receive impulses to and from other neurons. The space (or 'junction') through which a pair of adjacent neurons are able to communicate is referred to as a synapse. This communication is achieved by the release of neurotransmitters stored in vesicles that traverse from the neuron membrane at the presynaptic side to the receptors located on an adjacent neuron's membrane at the postsynaptic side. Synaptic connections between neurons allowing complex network behavior are known to play an important role in learning and memory. Therefore ,the ability of EM to obtain images with resolution high enough to observe synapses makes it an invaluable tool for neuroscience. Furthermore, EM has also been used to understand the morphology of cell types in different brain areas in different organisms. It also allows us to observe other neuroanatomical structures such as mitochondria, synaptic vesicles and microtubules.
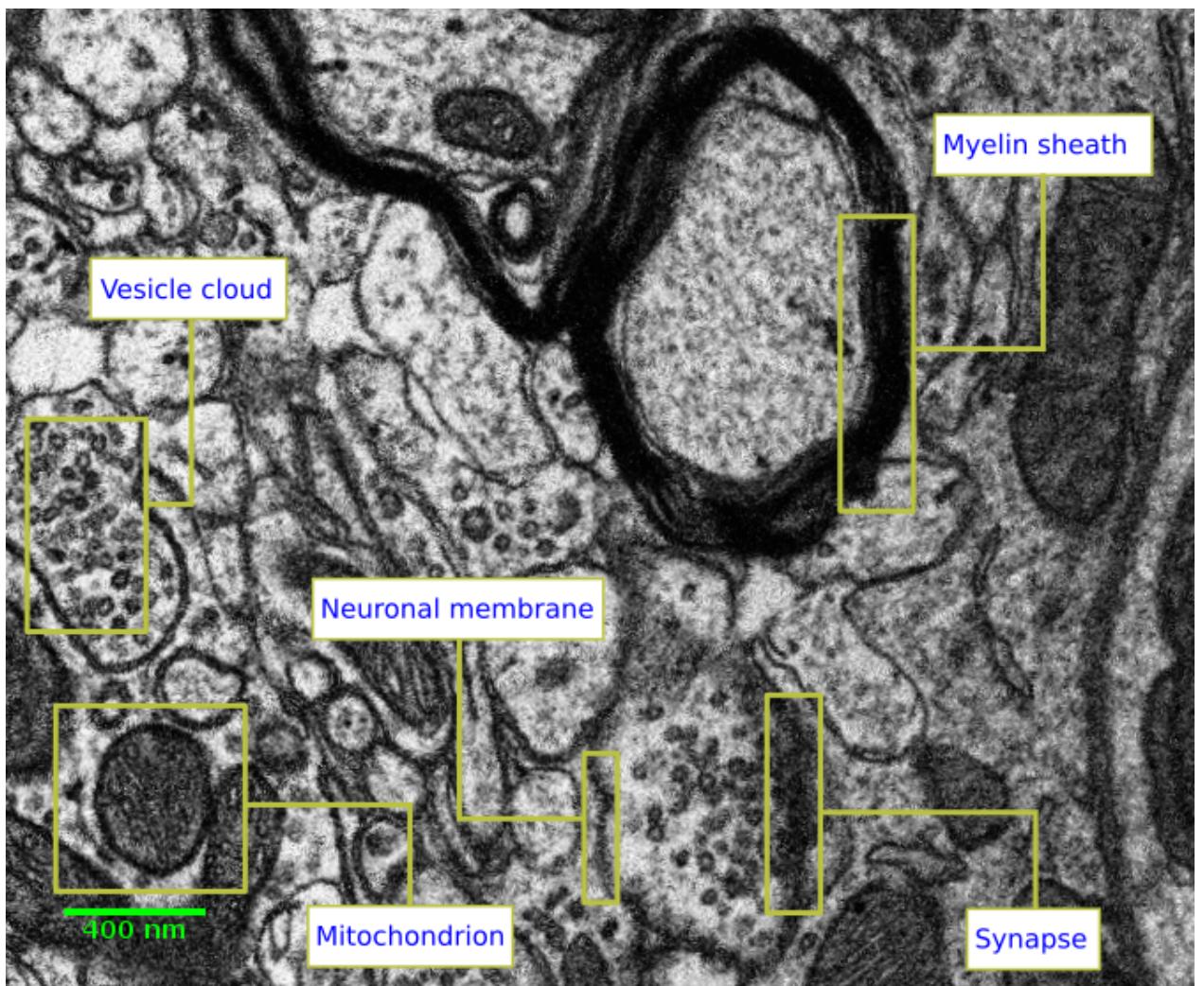
Figure 1.1: Part of an ssSEM image of a section obtained from zebra finch HVC. Image size: $550 \times 670\ pixels$ with pixel size $5\,\text{nm} \times 5\,\text{nm}$. *(Tissue section for imaging was provided by Ziqiang Huang.)*

1.1 Workflow: EM-image-driven neuroscience research

Volumetric EM images of a nervous system typically go through a sequence of steps. Once the biological tissue to be imaged is extracted, it has to go through a sample preparation protocol. This step ensures the sample is stable and ready to be cut in this sections. Once the sample is cut in thin sections a staining protocol is followed so that the sample produces an informative image subject to an electron beam. Depending on the type of EM, the thin sections have to be laid out on a suitable platform. Modern EM equipment allow automatic image acquisition for hundreds of sections positioned in series. Usually an individual section is too large to be imaged at one go. In such cases overlapping image patches are acquired and have to be stitched together using image processing tools. For serial section volumetric EM images, in addition to stitching them to form an image per individual section, they also have to be aligned across the Z-axis (orthogonal to the direction of cutting). To compensate different exposure levels occurring in different parts of the image, it is important to perform brightness and contrast adjustments. In situations where the exact volume corresponding to each section is of interest, section thickness estimation methods are deployed. Thus obtained EM images are ready to be annotated and analyzed.

1.2 Sample preparation

In order to examine biological tissue using EM, it has to be converted into a dry and stable material that interacts with an electron beam giving rise to the desired images. For ssTEM and ssSEM imaging in a vacuum, the sample preparation process typically involves the following steps [19, 26]:

- **Fixation** Biological tissue has to be transformed into a stable and preserved state by cross-linking macromolecules with the use of aldehydes.

- **Dehydration** In order to prevent any damage caused to the EM equipment when operating in a vacuum, the specimen has to be dehydrated. This is done by gradually substituting water in the specimen with a solvent such as ethanol or acetone. Dehydration also helps to preserve the specimen.

- **Embedding** The specimen needs to have high structural integrity and mechanical strength if it has to be sectioned or etched or fractured. To achieve this hydrophobic epoxy resins can be infused to the specimen followed by polymerization.

- **Thin sectioning** Serial section electron microscopy (ssEM) is the standard approach of obtaining volumetric image stack of a specimen which has dimensions of around tens of micrometers. This technique involves cutting the specimen into thin slices using a thin blade mounted on an equipment referred to as a microtome. Typical thickness of such a section varies from 30 nm to 80 nm. Sectioning is known to cause squashing of the specimen. Though occasional, it may also cause torn or missing sections.

- **Staining** EM image formation is based on the interaction between the specimen and a high energy electron beam. Therefore, the contrast between different struc-

tures (e.g. neuron membrane vs. cell interior) depends on how differently they interact with the electron beam. This step involves introducing heavy metal ions (e.g. lead, uranium) that preferably bind to different cellular structures which results in different contrast levels.

1.3 EM imaging

The operating principle of EM has some similarities to light microscopy (LM) except that instead of a beam of light EM uses a beam of electrons along with matching components to handle (e.g. focus and detect) electrons. The resolution of the images taken by LM is limited by the wavelength of light. Since EM uses electrons with much smaller wavelength as the source of light, it is able to image at nanometer resolution.

Serial Section Transmission Electron Microscopy (ssTEM)

An electron source emits an electron beam that travels along a vacuum guided and focused by electromagnetic lenses and goes through a thin specimen and reaches a detector which captures the image. Depending on the type and density of material present in the specimen some electrons are scattered away from the beam leaving a subset of the original beam reaching the detector. Different parts of the image appear with varying intensity according to the electron scattering occurred at the corresponding points of the specimen.

In ssTEM [35], several sections which usually correspond to a series of adjacent sections obtained from a sample are automatically imaged using the above technique. Therefore it's a common high throughput image acquisition method for obtaining volumetric image data of brain tissue. Each section typically can have up to a few hundreds of micrometers in length and width. These sections have to be thin enough to transmit a sufficient number of electrons to form an image. For sections obtained from brain tissue the thickness typically ranges from 40 nm to 70 nm.

Serial Section Scanning Electron Microscopy (ssSEM)

In SEM, a high energy electron beam interacts with a solid sample. This electron beam raster scans the surface of the specimen. The interaction between the specimen and the high energy electron beam gives rise to backscattered electrons and secondary electrons which are received at separate detectors. The signals generated by the inputs to the detectors are then used to create an image of the scanned surface. Similar to ssTEM, currently available ssSEM [38] equipment allows for high throughput imaging where a series of thin sections can be automatically scanned. The thickness of the sections used in ssSEM can be larger than in ssTEM because the beam does not have to be transmitted through the entire specimen. However in order to obtain a good SNR and to retain as much information as possible along the direction of cutting, the typical thickness of sections are similar in magnitude to those used in ssTEM.

Focus Ion Beam Scanning Electron Microscopy (FIB-SEM)

This imaging technique combines serial etching of a specimen block with scanning electron microscopy [49]. Etching the surface of a specimen is performed using a focused beam of ions that can be controlled with nanometer precision. Therefore, when combined with SEM, ion beam milling can be used to acquire volumetric EM images. The spatial interval between two adjacent surfaces of which images are acquired typically ranges from 5 nm to 10 nm which results in nearly isotropic volumetric images.

The main advantage of this technique is the high resolution obtainable along the axis perpendicular to the cutting plane compared to ssTEM and ssSEM. This is because the milling performed by a focus ion beam allows to etch thinner slices of specimen than what can be sectioned using a diamond knife of an ultramicrotome.

However, the size of the specimen that can be used in FIB-SEM is typically less than 10 µm in each dimension. This is much smaller compared to ssTEM and ssSEM. Therefore it's less suitable for volumetric imaging of whole brains than the other two.

1.4 Image registration

Automated electron microscopes allows us to image hundreds of large tissue sections. Large sections are usually imaged as a set of smaller tiles that have overlapping boundaries with their neighbours. An EM image of an entire 2D section has to be obtained by aligning and stitching together these image tiles. To reconstruct an accurate 3D EM image of the tissue, these 2D sections have to be aligned with respect to each other. Apart from alignment, image registration process has to also correct for nonlinear deformations of image tiles that might be caused during sectioning and imaging. The state of the art registration algorithm proposed by Saalfeld *et. al.* [68] jointly performs 2D stitching, 3D alignment and correction for deformations. Their approach is based on an initial alignment obtained by matching image landmarks of near by sections. These landmarks are defined using SIFT image feature extraction method [2]. Further deformations are estimated using local block matching. Afterwards, this initial alignment is optimized by modelling each section as a mesh of springs where parts of the image are allowed to translate and rotate subject to rigidity limits imposed by springs attached to it.

1.5 Image analysis in neuroscience

Methods and tools used for analysing EM images can be broadly categorized as follows:

(a) Post-acquisition image processing methods

(b) Visualization tools

(c) Annotation tools

(d) Semi-automatic segmentation methods

(e) Automatic segmentation methods

(f) Proofreading tools

Image postprocessing is required to compensate for deformations and distortions, stitch images together to form larger sections and align serial sections to form 3D image volumes. Once post-processing is done, the images are ready to be visualized and annotated.

Annotation of neuroanatomical structures and neurons in EM images are extensively used in anatomical and functional studies. Some of those studies involve quantitative comparisons of structural changes in the brain between groups of animals that can be subject to different experimental conditions. Such experiments are conducted to investigate the underlying mechanisms of learning and memory.

Neurons communicating with other neurons via synaptic junctions form networks of neurons that result in complex behaviors. There have been several attempts to build 3D reconstructions of such neuronal circuits from EM images of the brain during the last couple of decades. Neuronal circuit reconstruction mainly facilitates investigations into relationships between network structure and function, apart from anatomical properties of the networks themselves. Such studies have coined a subfield within neuroscience referred to as 'connectomics'.

Connectomics research depends on large EM datasets from which 3D neuronal circuits are derived. To this date most of such efforts are performed manually which is an extremely tedious process. When connectivity is the main concern, it suffices to annotate and connect the center of each neuron in a stack of serial section EM images. This is referred to as skeleton tracing and it's less tedious in comparison to annotating neuron boundaries. CATMAID [67] and KNOSSOS [36] are manual annotation tools that are being used to trace neuron skeletons in large datasets. However, to obtain a morphologically accurate representation of each neuron, the boundary of each neuron has to be outlined in each EM section. TrakEM2 [10] is a tool that can be used for this.

The vast amount of data that have to be annotated in order to reconstruct neural circuits require months of work from hundreds of people. On one hand researchers use crowd sourcing methods to scale up manual annotation efforts. Eyewire and Ariadne are examples of such efforts. The other way to make the annotation process more efficient is to use methods from machine learning and computer vision to achieve some level of automation. Most of the methods that can be used for automatic neuron segmentation rely on pixelwise probabilities obtained from learned classifiers.

Table 1.1: Image analysis, segmentation and annotation methods for EM images of brain tissue

Name of method	Description	Level of automation
CATMAID [67]	Neuron skeleton tracing	manual
Elegance [86]	Neuron skeleton tracing	manual
KNOSSOS [36]	Neuron skeleton tracing	manual
SSECRETT [41]	Neuron skeleton tracing	manual
Reconstruct [17]	Neuron morphology annotation (3D)	manual
TrakEM2 [10]	Neuron morphology annotation (3D)	manual
ITK-SNAP [87]	Neuron morphology segmentation (3D)	semi-automatic
NeuroTrace [42]	Neuron morphology segmentation using (3D)	semi-automatic
Icon [27]	Neuron morphology segmentation (3D)	semi-automatic
ILASTIK [73]	Pixel classification and clustering. Segmentation of synapses, membranes and neurons.	semi-automatic
GALA [58]	Dense neuron segmentation (2D and 3D) using superpixel agglomeration	automatic
Segmentation Fusion [82]	Dense neuron reconstruction	automatic
SOPNET [23]	Dense neuron reconstruction (3D)	automatic
RhoANA [77]	Dense neuron reconstruction (3D)	automatic

1.6 Thesis organization

This thesis is organized in two main parts. The first part containing chapters two and three focuses on image processing and computer vision method development. Chapter 2 presents a method that can be used to estimate the XY anisotropy and Z thickness of a serial section EM image stack. It also provides details of a validation dataset we created using atomic force microscopy which we used to compare the performance of our method with other methods. Chapter 3 describes a method to segment 2D neuron slices on individual EM sections using topological constraints. According to the results obtained, this approach lowers the number of segmentation errors compared to other 2D segmentation methods.

The second part of the thesis provides an image driven analysis of a neuroanatomical changes caused by experience dependent learning. Chapter 4 provides an introduction to zebra finch vocal learning and Chapter 5 contains the experimental details and the analysis performed along with the results obtained. Chapter 6 is an overall discussion about the thesis including an outlook for the work carried out as part of this dissertation.

Part I

Methods for analysing electron microscopy images

Chapter 2

Thickness and anisotropy estimation of EM sections

2.1 Introduction

Electron microscopy (EM) has enabled imaging of nano-scale neuroanatomical structures such as synapses. Images acquired using EM have been instrumental in neuroscience research because of the level of information provided which is unmatched by any other imaging technique. Serial section Scanning Electron Microscopy (ssSEM) and serial section Transmission Electron Microscopy (ssTEM) are used to inspect tissue volumes on the scale of tens to hundreds of micrometers in each dimension. Tissue sections suitable for ssEM typically have a thickness that ranges from 30 nm to 70 nm. These thin serial sections are cut from a specimen embedded in resin using a microtome which is equipped with a diamond knife. It has been quantified that the actual thickness of the sections obtained can be different from the expected thickness by up to 20% [20]. Another EM technique used to obtain volumetric image data is Focused Ion Beam Scanning Electron Microscopy (FIBSEM) which allows much finer sectioning (e.g. 5 nm $\sim$ 10 nm) compared to what is currently achievable with ultramicrotomes. The problem of section thickness variation is also observed in FIBSEM data [49].

Image processing methods commonly implicitly assume isotropy of physical structures along the imaging plane [24]. However, sources of anisotropy (stretching) in the imaging plane include anisotropy intrinsic to the specimen, effects of sample handling and cutting, and imperfections in microscope calibration. We focus on the image analysis problem of determining the overall stretching, without distinguishing between the individual sources of stretching.

In this work we address estimation of thickness and stretching by learning a function f that captures the variation of image statistics. This function can be used to infer the spatial distance along the Z axis between pairs of images in an image stack, and the anisotropy of the imaging plane XY. To compute predictive distributions of spatial distance for new, unseen images we use Gaussian Processes (GPs) [64] to perform a non-parametric Bayesian regression. We also use GP regressors to estimate stretching.

Section thickness estimates can be used to correct volume estimates along the Z axis (perpendicular to the cutting plane), which is useful for producing more accurate 3D

reconstructions of imaged tissue. Furthermore, the accuracy of counts of objects such as synapses normalized per unit volume can be improved using both the thickness estimates and the stretching estimates. For instance, accurate section thickness estimates can be used when applying the disector method [75] which is one of the most commonly used methods to get an unbiased estimate of objects (eg. synapses) within a 3D volume. Disector method essentially quantifies the number of objects of interest in a virtual 3D volume by sampling a few consecutive 2D sections of the volume. Correct estimates of section thickness is highly beneficial in such cases to improve the accuracy of the quantifications thus obtained.

For any method that uses the known XY resolution to model the absolute spatial distance between sections (including [74] and our method), it is important to have an estimate of the anisotropy along the XY plane. Such XY anisotropy lead to different statistics along the two axes. If unaccounted for, the disparity of these statistics will introduce inaccuracies in the results obtained by methods that readily assume similar statistics along the two axes. Our solution to this problem is described in Section 2.8.

In order to validate the thickness estimates, we have directly measured the thickness of a set of EM sections using atomic force microscopy (AFM). We have made the validation dataset publicly available as a benchmark to evaluate section thickness estimation methods.[1] Validation results and estimates for z-section thickness and xy-anisotropy for FIBSEM, ssTEM and ssSEM data sets are discussed in Section 2.10.

Publication and source code

The method proposed in this chapter is intended to be published in an article titled "Estimation of Z-thickness and XY-anisotropy of electron microscopy images using Gaussian Processes" (Ambegoda et al.). The code for running the experiments described in this chapter is publicly available.[2]

---

[1]https://github.com/thanujadax/ssSEM_AFM_thickness
[2]https://github.com/thanujadax/gpthickness
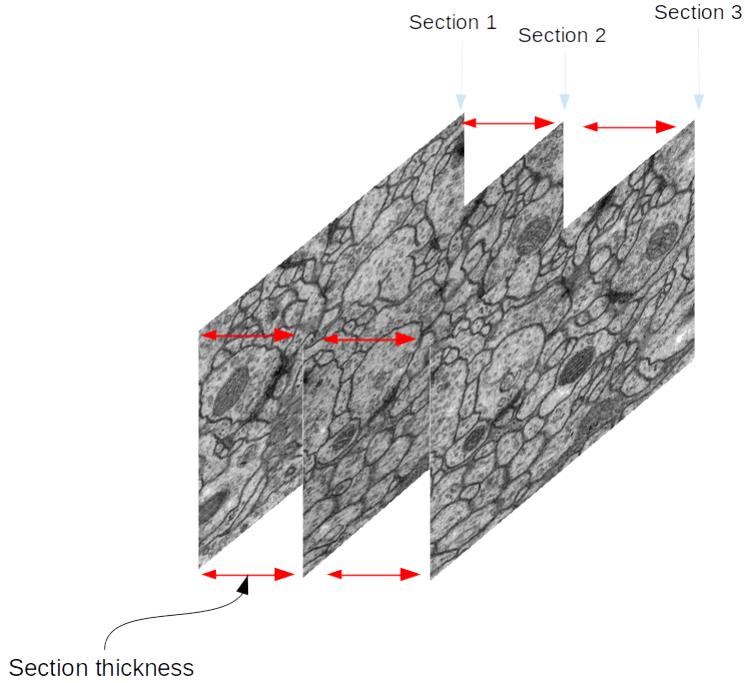
Figure 2.1: In a given ordered series of EM images, we assume that the distance between the section i and section $i+1$ is the same as the *thickness* of section i.

2.2 Related work

Methods used to indirectly estimate EM section thickness can be broadly categorized into three types:

1. Thickness estimation using ellipsometry

   The earliest thickness estimation method suggested in [61] uses the reflected color of the thin sections floating on water to coarsely estimate the thickness of the sections. Here, thickness estimation is done by matching the color of the sections with a predetermined color code that relates the colors to different thickness values. This method provides a coarse estimate with an accuracy around 30 nm. The minimum thickness that can be estimated is about 50 nm. This method is still commonly used to get a first idea of the thickness of sections being cut.

2. Thickness estimates by detecting special structures within the images

   In [72], accidental folds of the thin tissue section seen in the electron micrograph are used to determine the thickness of that section. Cylindrical diameters method was proposed in [18] where use nearly "cylindrical" mitochondria to get an estimate of section thickness under the assumption that we are able to find cylindrical mitochondria that exist vertically parallel to the cutting direction.

3. Computational approaches using image statistics

   In [74] the relationship between pairwise image dissimilarity and distance is computed as the average of discrete data points and the estimated thickness of a new section is interpolated from these. Both [74, 31] assume that locally, images are realizations of an isotropic and rotationally invariant process. By contrast, we adopt an approach that is less affected by sample anisotropy in the xy plane. In [31] the positions of the images along the Z axis are iteratively corrected to seek a consistent solution in which adjacent sections have an optimal gap (or thickness) between each other. The optimal solution adjusts the positions of the images such that the distance-similarity curve is maximally smooth after a fixed number of iterations. The method presented in this chapter falls into this category as well.

2.3 Contributions

Ziqiang Huang and Thomas Templier from the Institute of Neuroinformatics, University of Zurich and ETH Zurich provided samples used for microscopy. FIBSEM images and some of the ssSEM images were provided by Ziqiang Huang. Atomic Force Microscopy was performed by Jozef Adamcik of the Department of Health Science and Technology, ETH Zurich. Gaussian Process modeling was done by Julien Martel of the Institute of Neuroinformatics, University of Zurich and ETH Zurich.

2.4 Image similarity

An image stack obtained from ssEM corresponds to consecutive sections obtained from a 3D volume which in our case is typically a part of brain tissue prepared for EM imaging. As illustrated in Figure 2.2, images which are close to each other in the volume tend to be more similar to each other than the ones which are further apart. This is because the anatomical structures that have been imaged vary their visual properties smoothly in the 3-dimensional space compared to the resolution of the imaging. This correlation between the similarity of a pair of sections and the distance between them has been used in [74] to build a predictive function to estimate section thickness. Our approach is also based on the relationship between image similarity and the distance between them. We differ to [74] in the way we learn the predictive function.

<div align="center">(a)　　　　　　　　　(b)　　　　　　　　　(c)　　　　　　　　　(d)</div>
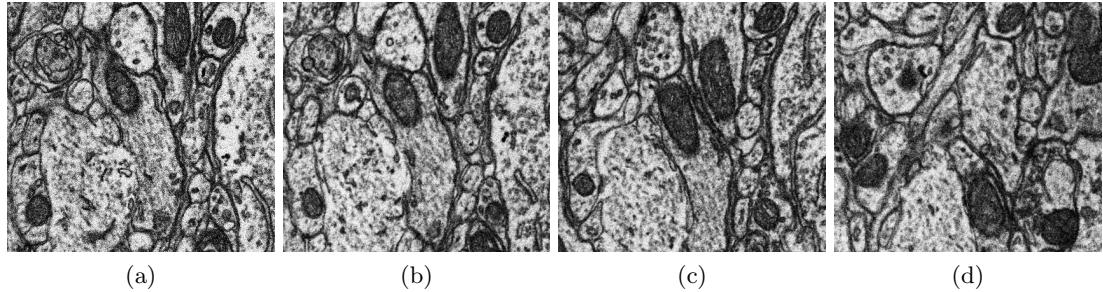
Figure 2.2: Images taken from a single stack of ssSEM images at different positions of the Z axis where Z axis of the image stack is perpendicular to the cutting direction. The expected distance between two sections at adjacent Z positions is approximately 70 nm. Each image corresponds to an area of $2\,\mu m \times 2\,\mu m$. (a) Image at position $z = 1$. (b) Image at position $z = 2$. (c) Image at position $z = 5$. (d) Image at position $z = 15$. It can be observed that the biological structures such as neuron membranes, mitochondria, synapses and vesicles displayed in the images are similar when the images are close to each other. When the distance between two sections increase there is less similarity between them.

2.5   Preprocessing: contrast adjustment and registration

Raw images obtained from EM might have different exposure levels at different parts of the image, even when the image is a single tile as shown in Figure 2.3a. The general approaches to contrast enhancement is based on stretching the distribution of the pixel intensity values so that there is enough difference of intensity between regions that should show high contrast with respect to each other. One such common approach is histogram equalization. Here an optimal transformation is applied to the entire image to improve its contrast. When the original image contains regions of different contrast levels, this approach is not able to sufficiently enhance contrast. To overcome this problem, we use contrast limited adaptive histogram equalization (CLAHE) [91] to make the contrast level more homogeneous over the entire image as illustrated in Figure 2.3b. Adaptive histogram equalization (AHE) performs contrast adjustments for each pixel using a transformation function derived using the histogram around that pixel. Contrast limitation is used to reduce the amplification of noise while performing AHE.

After acquiring the images from a series of images from the same volume, the images have to be registered with each other. Registration process makes sure that each image is optimally aligned with each other along the entire stack of images. Normalizing the contrast levels and proper registration is a very important preprocessing step in any method that's based on calculating the similarity of images because registration problems and contrast variations will introduce an erroneous dissimilarity in addition to the expected dissimilarity that is caused by the distance between the sections.

<div align="center">(a)   (b)</div>
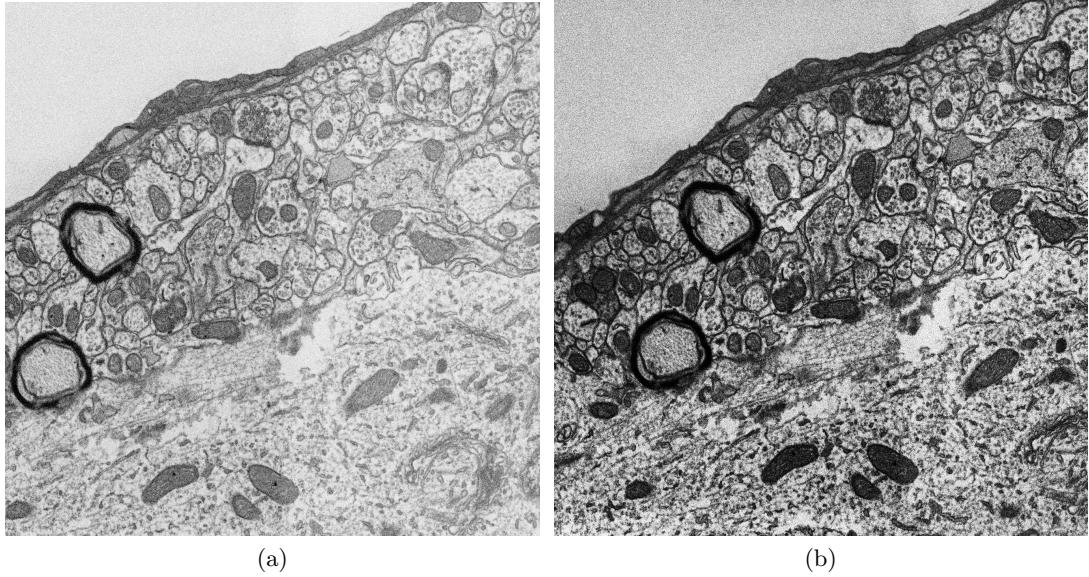
Figure 2.3: (a) Raw ssSEM image with varying contrast at different locations. The top left quarter of the image has higher contrast compared to rest of the image. The bottom right quarter seems to be over exposed. (b) After performing contrast limited adaptive histogram equalization [91]. Both images are of size 1350×1350 pixels with each pixel of size $5\,\text{nm} \times 5\,\text{nm}$.

2.6   Estimation of section thickness

We propose to learn a function of pairwise image dissimilarity to estimate the distance between the pair of sections. Our approach adapts the work of [74] with the variation described below. What we refer to as *section thickness* is the distance between a pair of adjacent sections in a series of volumetric images (Figure 2.1).

We assume that local structures in the images vary smoothly in all directions at a spatial scale larger than the section thickness. Hence, the dissimilarity S_{I_A, I_B} between two parallel images I_A and I_B only depends on the spatial distance $D_{A,B}$ between them. To learn the variation of image dissimilarity as a function of the distance we first have to obtain distance-dissimilarity data points. This can be done by extracting images at known different distances along the X and Y axes of the imaging plane: $D_{A,B} = f(S_{I_A, I_B})$. From an original image I, two image patches A and B which are a distance $D_{A,B} = n \times \Delta x$ away from each other can be generated by extracting image patch A centered on pixel coordinates (x_i, y_i) (on image I) and an equally sized image patch B centered on pixel coordinates $(x_i + n \times \Delta x, y_i)$ on image I. Here Δx is the length of a rectangular pixel along the x axis and n is the number of pixels.

As the dissimilarity measure S_{I_A, I_B} we use the standard deviation of pixel-wise intensity differences (SDI) defined in Equation (2.1), similar to [74].

$$S_{I_A, I_B} = \sqrt{\frac{1}{N} \sum_{x,y} (I_{A_{x,y}} - I_{B_{x,y}})^2} \tag{2.1}$$

We learn two separate distance-dissimilarity functions $f_x(S)$ and $f_y(S)$ as described

in Section 2.7 using images generated along the two axes X and Y. After estimating the relative stretching γ between the two axes (Section 2.8), we use one of these functions to estimate section thickness depending on the value of γ. For most accurate results, the regressor should be the one that has been learned along the direction of minimal stretching. Since we are generally pointed to a possibility of samples being compressed in one direction relative to the other due to effects of tissue embedding/handling/cutting we recommend using the function corresponding to the lesser compressed axis. Determination of the direction along which maximum stretching is described in Section 2.8.

2.7 Distance-Dissimilarity function using Gaussian process regression

The function $f(S)$ we aim to learn takes the dissimilarity S of a pair of images and returns the distance D between those two images:

$$D = f(S) \tag{2.2}$$

In order to learn such a function, a training data set in the form of distance-dissimilarity tuples has to be created. Given a set of images we can create such a data set of N data points $\{(D_i, S_i)\}_{i \in \{1,\dots,N\}}$ using the sliding window approach suggested above. Such a set of tuples is generally referred to as labeled training data. We can use these data points to perform regression which results in the desired function $f(S)$. Once $f(S)$ is learned, it can be used to predict distances D between a new pairs of images for which the similarity can be directly calculated.

Regression

The general approach for learning a parametric regression function is to first assume a specific function $f(S; \mathbf{w})$ with a set of parameters w to be determined using training data. Therefore, the regression problem essentially becomes an optimization problem that learns the best values for the set of parameters. This is done by minimizing a sample loss L_i for each training data point that provide input values S_i and output values D_i which are to be mapped by $f(S; \mathbf{w})$. For instance, in least squares regression the following optimization problems gives the optimal parameter vector $\mathbf{w}^*$:

$$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} \ \sum_i \|D_i - f(S_i; \mathbf{w})\|_2^2. \tag{2.3}$$

An alternative formulation of the regression problem is provided within a Bayesian framework. The regression task then becomes equivalent to inferring the posterior distribution $p(\mathbf{w}|D, S)$ of the parameters given the training data, the prior distribution of the parameters $p(\mathbf{w})$ and the likelihood $p(D|S, \mathbf{w})$. Here, the following underlying model can be assumed to give rise to the data D, S:

$$D = f(S; \mathbf{w}) + \epsilon, \tag{2.4}$$

where ϵ represents the noise that explains any deviations of D from the mapping $f(S, \mathbf{w})$.

The mostly likely regressor is then given by the mode of the posterior $p(\mathbf{w}|\{D_i, S_i\}_i)$ and the uncertainty of the prediction will be given by the standard deviation of the posterior.

An obvious limitation of a parametric regression is the need to have $f(S; \mathbf{w})$ explicitly specified. In general for practical problems this the functional form of f not available *a priori*. Therefore, any formulation of $f(S; \mathbf{w})$ in a parametric setting would be influenced by strong underlying assumptions about the process that gives rise to the data.

Gaussian Process regression

Our regression problem can be formulated in function space [64] using Gaussian process regression [84]. A Gaussian process (GP) can be defined by a set of random variables for which any finite subset has a joint Gaussian distribution. A GP is fully determined by a mean function $m(S)$ that provides information about the mean values:

$$m(S) = \mathbb{E}\big[f(S)\big], \tag{2.5}$$

and a covariance function $k(S, S')$ that captures the covariance of the process $f(S)$:

$$k(S, S') = \mathbb{E}\big[\big(f(S) - m(S)\big)\big(f(S') - m(S')\big)\big], \tag{2.6}$$

where $\mathbb{E}$ denotes the expectation. The unknown regressor $f(S)$ is then a realization of the following GP:

$$f(S) \sim \mathcal{GP}\big(m(S), k(S, S')\big). \tag{2.7}$$

In order to perform Gaussian process regression [85], only a family of functions for m and k have to be specified. These functions do not enforce a form for f. The covariance function k specifies a prior over the space of functions. Furthermore, hyperparameters for the mean and covariance functions can be learned from data.

For our model, we choose the covariance $k(S, S')$ function to be a squared exponential (SE):

$$k(S, S') = \sigma^2 \exp\Big(-\frac{1}{2\,l^2}\,(S - S')^2\Big) \tag{2.8}$$

SE based covariance function is a smoothness prior on the functions determined by the length-scale l and signal standard deviation parameter σ. For the mean function a function in the form $m(S) = a\,S^b$ was chosen for the distance measure SDI which empirically behaves as a power law increase starting from $(0, 0)$.

$$m(S) = a\,S^b \tag{2.9}$$

For a distance measure that starts from a constant value at zero distance (e.g. Pearson correlation coefficient) and drops with increasing S according to a power law, a function in the form given by equation (2.10) can be used.

$$m(S) = c - a\,S^b \tag{2.10}$$

Hyperparameters $\theta = (\sigma, l, a, b)^T$ are learned from the training data by optimizing the marginal likelihood using non-flat hyperpriors over the mean hyperparameters.
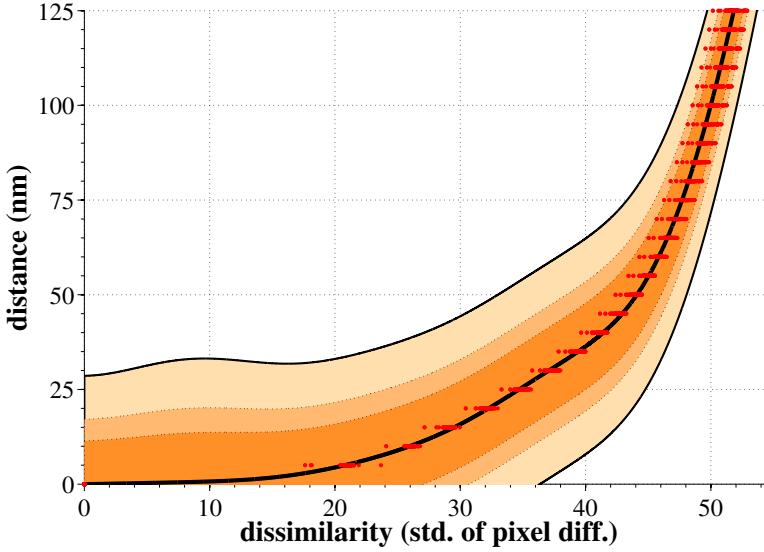
Figure 2.4: Graph of distance D vs. image dissimilarity s, $D = f(S)$, used for the estimation of section thickness and stretching. Shown are the training data (blue dots), the mean (bold line), and multiple standard deviations σ of the Gaussian Process (GP) predictive distributions (darkest to lightest orange) $2\,\sigma$(95%), $3\,\sigma$(99.7%), $5\,\sigma$(99.$\underline{9}_4$%).

These hyperpriors are only used to initialize the optimization but are not restrictive. Both hyperpriors $p(a) = \mathcal{N}(\mu_a, \sigma_a)$ and $p(b) = \mathcal{N}(\mu_b, \sigma_b)$ are Gaussian distributions whose means (μ_a, μ_b) and standard deviations (σ_a, σ_b) are determined using a standard nonlinear regression of a function $D' = a \cdot S'^b$ using a Levenberg-Marquardt algorithm.

To predict unknown D values for given S values for new pairs of images by performing GP regression, the marginal likelihood of the GP has to be computed. Then a predictive distribution is produced for output D at each test input S. Figure 2.4 shows an instance of the mean and variance of these predictive distributions plotted as a distance-dissimilarity curve.

2.8 Estimation of stretching

The learned distance-dissimilarity function can be utilized to estimate the *stretching coefficient* γ, defined as the deviation from isotropy of the image along one axis relative to the other. Consider a small image patch with pixel intensity gradient $\vec{g}$ at an angle θ relative to the x axis (Figure 2.5). The intensity difference Δp_i at pixel i between two image patches separated by one pixel (Δx) along the x axis is given by: $\Delta p = p_i - p_i \|\vec{g}\| \cos\theta \Delta x$, where p_i is the pixel intensity at pixel i. It follows from Equation (2.1) that the dissimilarity between these image patches (ignoring boundary conditions) is:

$$S(I_{x,y}, I_{x+\Delta x,y}) = \Delta x \, \|\vec{g}\| \cos\theta \sqrt{\frac{1}{N}\sum_{i=1}^{N} p_i^2} \qquad (2.11)$$
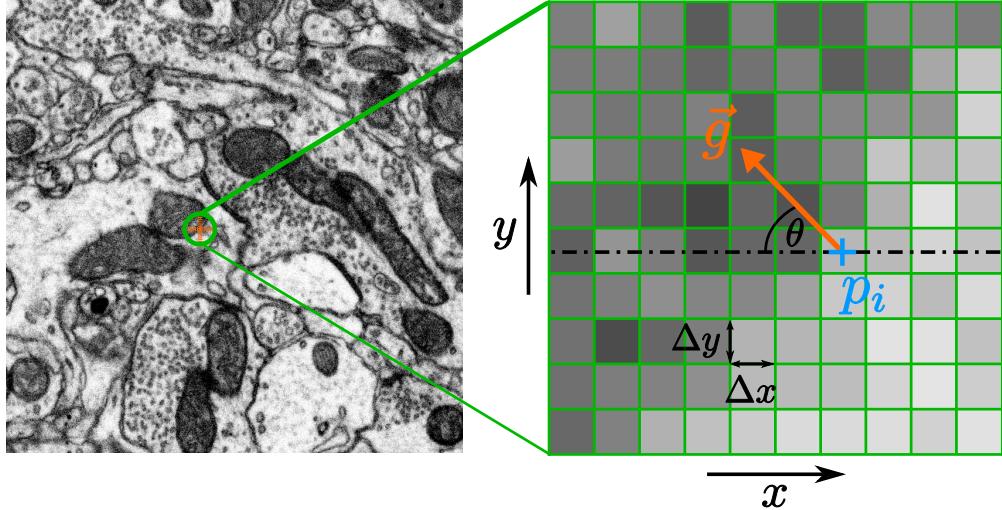
Figure 2.5: Left: FIBSEM image of 700×700 pixels. Right: An image patch with gradient $\vec{g}$ located at p_i with an angle θ w.r.t to the x axis.

As shown by Equation (2.11) the dissimilarity is directly proportional to the local gradient of the image patch. We use this result to estimate γ along one axis relative to the other (because stretching along one axis alters the component of the gradient along that axis).

To estimate γ along the y axis relative to the x axis (i.e. γ_{yx}) we perform the following steps: First we learn the distance-dissimilarity function $f_x(S)$. To do so, pairs of images that are displaced from one another by n pixels along the x axis can be used. Then, for one pair of images separated by one pixel along y axis (distance Δy), the distance can be estimated using $f_x(S)$. This estimate gives $\hat{n}_{yx}$, where $\hat{\Delta y} = \hat{n}_{yx}\, \Delta x$. This is the expected length of a pixel along y axis using the distance-dissimilarity statistics along x axis. Therefore, $\hat{n}_{yx}$ captures the linear scaling of y axis with respect to x axis in terms of distance-dissimilarity statistics. The stretching coefficient γ_{yx} of y axis relative to x axis is defined as

$$\gamma_{yx} = \frac{\Delta y}{\hat{\Delta y}} = \frac{\Delta y}{\hat{n}_{yx}\, \Delta x} = \frac{a_{yx}}{\hat{n}_{yx}} \tag{2.12}$$

where a_{yx} is the pixel aspect ratio $\Delta y / \Delta x$. For a pixel aspect ratio of 1, $\gamma_{yx} > 1$ implies stretching of the y axis relative to the x axis. Once the γ_{yx} is known, we suggest to use the regressor corresponding to higher γ (lower relative compression) as the distance-dissimilarity function for section thickness estimation. For instance, provided $\gamma_{yx} < 1$, the regressor $f_x(S)$ should be used since the linear compression of y axis is potentially more than that along x axis and therefore $f_x(S)$ will result in a more accurate thickness estimate.

However, the exact orientation of the X and Y axes are arbitrary. In order to find the directions of maximum and minimum stretching, γ_{yx} for a range of orientations has to be calculated. The lowest value of γ_{yx}^* corresponds to the pair of orthogonal axes where X has the minimum stretching along its direction.

1×1 1×0.75 1×0.50 1×0.25

(a)

(b)   (c)
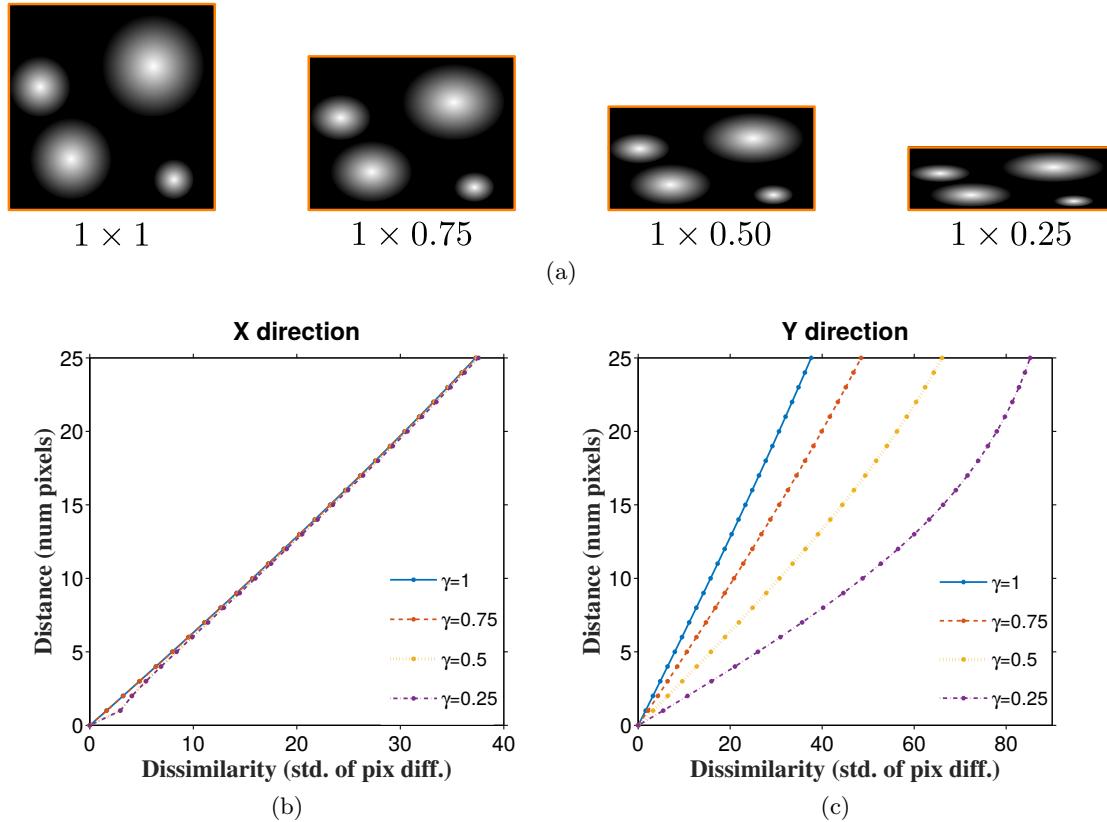
Figure 2.6: (a) shows four images with an artificially generated pattern featuring locally radial gradients. From left to right the images underwent a simulated linear stretching (in this case, compression) along the y axis while the x axis remained unchanged. The legend specifies the normalized size $w \times h$ of the image. (b) shows the distance-dissimilarity plots for shifted versions of the radial gradient pattern along the horizontal axis and (c) along the vertical axis.

2.9   Validation of thickness estimation using AFM

Validation of EM section thickness estimation methods is limited by the unavailability of a standard data set with accurately measured thickness. We use *Atomic Force Microscopy* (AFM) [6] to produce such a dataset for validation of thickness estimates using our method and to compare with results from other methods. AFM is a scanning probe microscopy technique that can be used to measure the 3D surface profile of a sample at nanometer resolution. The AFM probe which is a sharp tip with a typical radius of $5\,\mathrm{nm} \sim 50\,\mathrm{nm}$ scans over the surface while measuring changes in the atomic forces between the sample and the tip. This allows us to directly measure the thickness of ssEM sections which are placed on silicon wafers.

As shown in Fig. 2.8a thickness measurements were obtained using AFM along three distinct axes along each EM section. For each section, the average difference of these three height profiles (Fig. 2.8b) between the surface of the silicon wafer and the surface of the EM section is taken as the measured thickness.
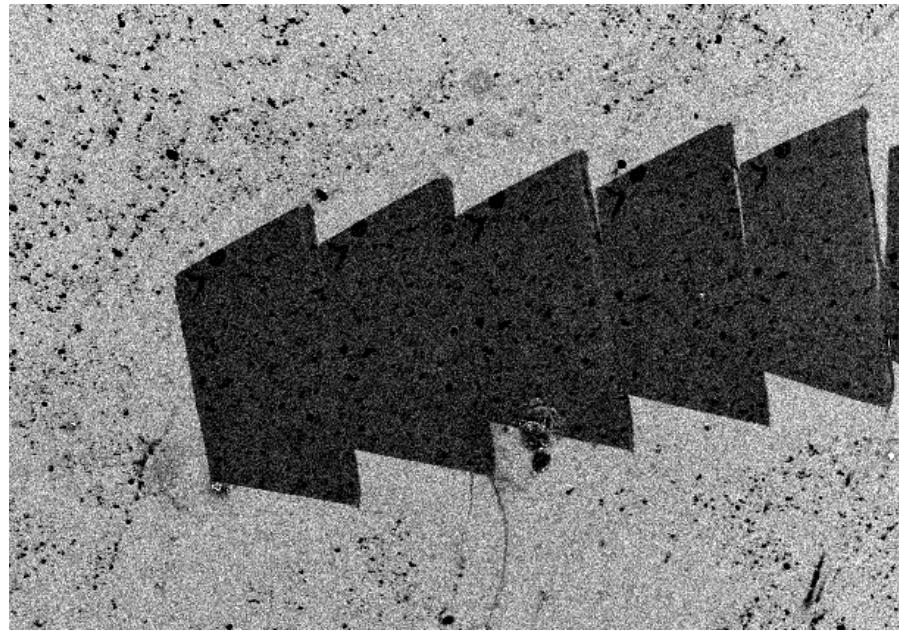
Figure 2.7: EM image of several trapezoid shaped EM sections placed on a silicon wafer. The image was obtained at a pixel resolution of $2.2\,\mu m \times 2.2\,\mu m$.
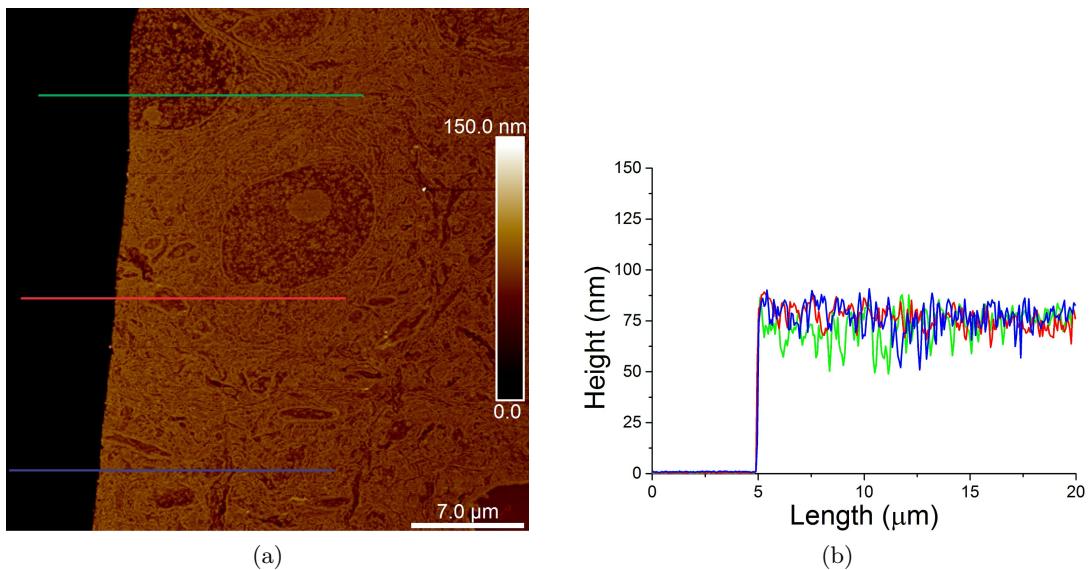
Figure 2.8: Figure 2.8a is an AFM image of a section prepared for EM. The pixel intensity corresponds to relative height difference from a reference point. Here the reference $-20\,nm$ corresponds to the level of the silicon wafer on which the sections are mounted. We set this level to zero in order to obtain the absolute thickness values. For each section we obtain measurements along 3 axes (green, red and blue) as shown in the image. (b) shows the measured height differences relative to the height of the silicon wafer along each of the 3 colored lines in Figure 2.8a.

2.10   Results and discussion

Anisotropy estimation

(a)

(b)
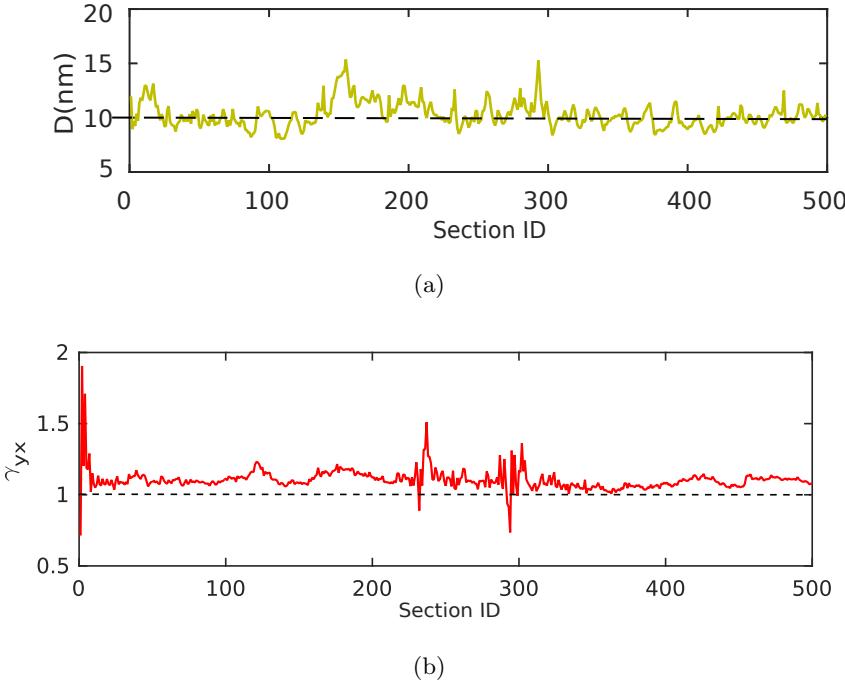
Figure 2.9: (a) Estimated section thickness (D) for a FIBSEM stack of 500 consecutive images (expected: $D \approx 10\,\text{nm}$) and (b) Estimates of the stretching coeff. γ_{yx} for a FIBSEM stack of 490 sections.

To validate the estimation of stretching coefficient γ we used different linearly stretched versions of a synthetic image as shown in Figure 2.6a. The original image was composed of bright circular objects with radial gradients. Then the image was re-scaled with known γ along the y axis (vertical) down to different sizes. Using equation (2.12) we recovered γ with an average accuracy of 97.3% for a linear compression of 75% of the original size (Table 2.2).

Estimated γ for real data sets ssTEM [1] and FIBSEM are summarized in Table 2.2. The FIBSEM dataset with 500 sections was from songbird brain tissue obtained at $5\,\text{nm} \times 5\,\text{nm}$ resolution on the xy plane with expected section thickness of $10\,\text{nm}$. The entire FIBSEM stack has the dimensions $8\,\mu\text{m} \times 8\,\mu\text{m} \times 5\,\mu\text{m}$. The ssTEM dataset of 20 serial sections was obtained from the same brain area. We note that FIBSEM shows a higher γ (corresponds to lower linear compression) compared to ssTEM. This correlates with the fact that unlike in ssTEM, FIBSEM doesn't make use of a diamond knife for thin sectioning which is a potential source of linear compression. Instead it uses an ion beam to successively burn away thin layers.

As mentioned in Section 2.8, the image axis X and Y are arbitrarily chosen. Therefore we estimate the maximum stretching factor γ_{yx}^* for a range of possible axes by rotating the original images up to $180°$. Anisotropy estimates for a range of such rotations is shown in Figure 2.11. For better thickness estimation accuracy, we argue

that the regressors obtained from the images that are rotated to give the maximum γ_{yx} have to be used. At this rotation the X axis is minimally stretched compared to the Y axis. This argument is more effective when the images are large enough to not have any intrinsic anisotropy towards a particular direction.

Thickness estimation

For the estimation of section thickness using the regressor f_x learned using GP has to be learned along the direction of least stretching. To select the f_x where the X axis is aligned with the least stretched direction, γ_{yx} has to be calculated for different rotations of the X and Y axis for the same image stack.

We prepared a dataset of 20 consecutive ssSEM sections of size $10\,\mu\text{m} \times 10\,\mu\text{m}$ from songbird brain tissue (HVC) to validate our thickness estimation method. EM images were obtained at a spatial resolution of $5\,\text{nm} \times 5\,\text{nm}$. The thickness of each section was measured using AFM as described earlier in Section 2.9. Three different image stacks of size 1300×1300 were prepared from the first image stack. The average thickness measured using AFM was $74.35 \pm 2.64\,\text{nm}$. Our method was able to estimate the thickness of the 20 sections with a mean absolute error of $9.91\% \pm 1.97$ whereas the XY-averaging method suggested in [74] estimated the same stack with a mean absolute error of $18.26\% \pm 1.04$. The full comparison of estimation errors is plotted in Figure 2.9 and Figure 2.12.

In addition to thickness measurements of ssSEM sections using AFM, we propose a second approach for validating thickness estimates using synthetic sections with known thickness derived from nearly isotropic FIBSEM volumes with known XY resolution. We use the method described in Section 2.6 to generate data points for learning function (2.2). For validation of estimated section thickness we split each image stack into separate training and test data sets. The training sets were used to learn the regression function given by Equation (2.2) as described in Section 2.7 and the test images were used for validation. We trained a regression function on 100 images of size 1400×900 pixels from a FIBSEM image stack, Figure 2.4. We used the test images to create 3 separate image sequences of 30 images each with known displacements of $10\,\text{nm}$, $50\,\text{nm}$ and $75\,\text{nm}$ along the relatively uncompressed axis. The results obtained are summarized in Table 2.1 along with a comparison with [74].

Although included for comparison, we note that in [74] an average distance-dissimilarity curve is generated for each pair of images between which the distance has to be estimated and therefore the interpolation function is based on the statistics of the validation data itself, unlike in our approach. Since in [31] only relative coordinates for each section were calculated we were not able to compare our numbers with this latter work.

The estimates of anisotropy and thickness from our method is sensitive to the image content and registration problems. When images happen to contain anisotropic patterns such as elongated mitochondria aligned in one direction more than others, it increases the stretching coefficient along that direction which causes the regressor f_x learned along the least stretched direction to be slightly inaccurate. This particularly takes place when the image size is not large enough to contain a large number of variably oriented structures so that there is no preferred orientation on average. When the regressor f_x is being used to estimate section thickness, any registration problems between a pair of adjacent sections

Table 2.1: Average thickness estimates for sets of 30 sections with the "ground truth" thickness generated as described in Section 2.6.
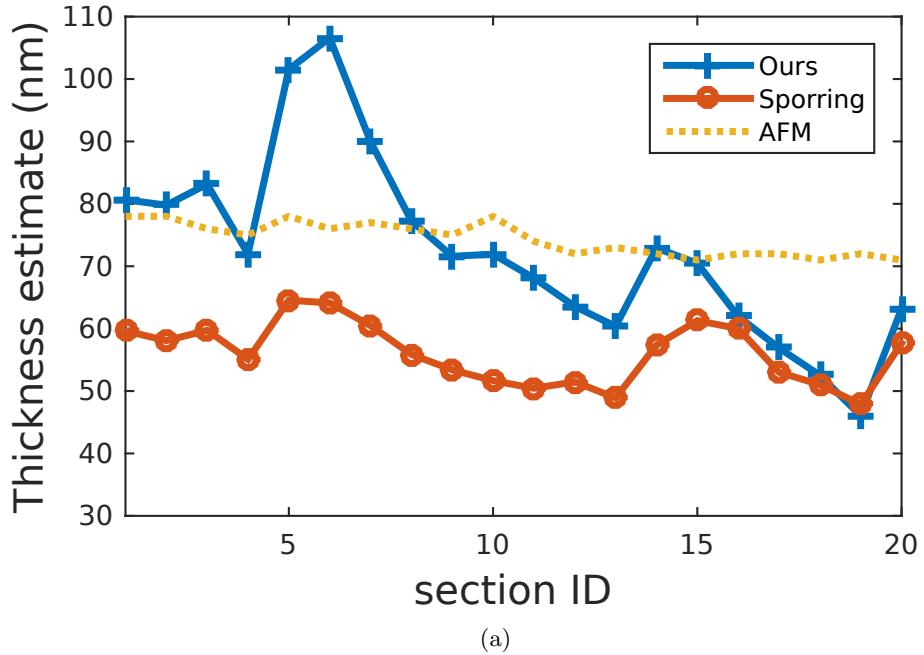
	Thickness values are in nanometers (nm)		
"Ground Truth"	**10**	**50**	**75**
xy avg. [74]	9.93	47.35	69.09
Ours	10.18 ± 5.61	47.02 ± 5.60	71.36 ± 5.59

Table 2.2: Estimated stretching coefficient γ_{yx} using synthetic images (Figure 2.6), 500 FIBSEM images and 20 ssSEM images.
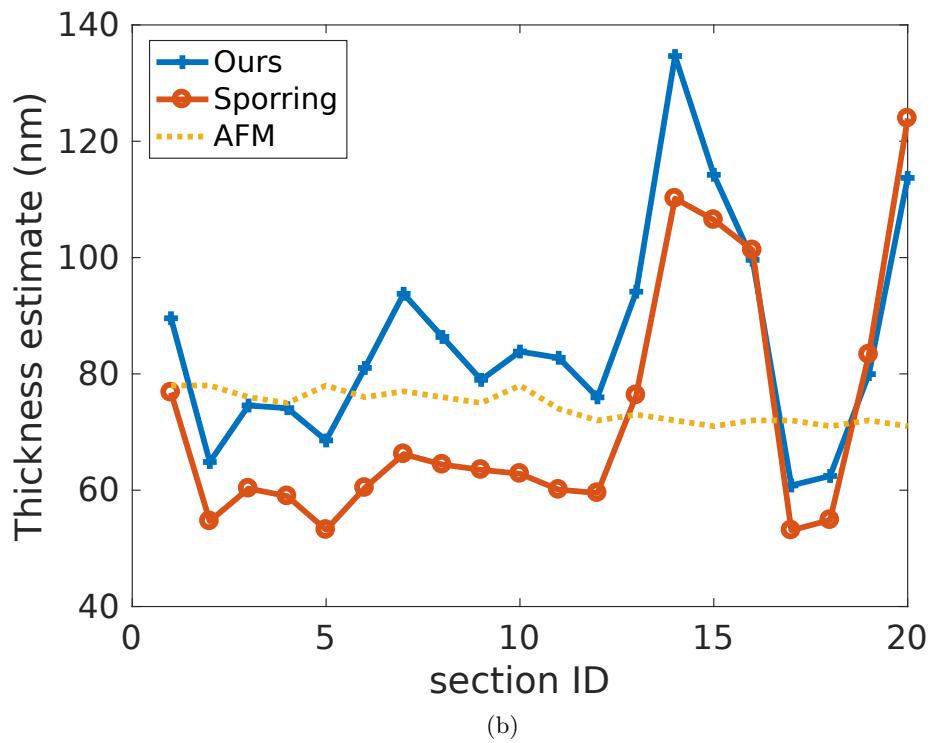
γ **for synthetic images (Figure 2.6)**			γ **for real images**	
Ground-Truth	*0.75*	*0.50*	FIBSEM	$0.94 \pm 6.6 \times 10^{-4}$
Estimates	0.73	0.63	ssSEM	0.86 ± 0.01

leads to an additional distance between them. Therefore it has to be made sure that the stack of images are properly registered before being used for thickness estimation. We observe that the registration error tends to increase with image size. Therefore it limits the maximum size of the images to be used for estimating thickness. We observe that images with 3000×3000 pixels where the pixel size is $5\,\text{nm} \times 5\,\text{nm}$ are large enough not to have a preferred orientation of structures on average, as well as small enough not to have a significant registration error.

(a)

(b)

(c)
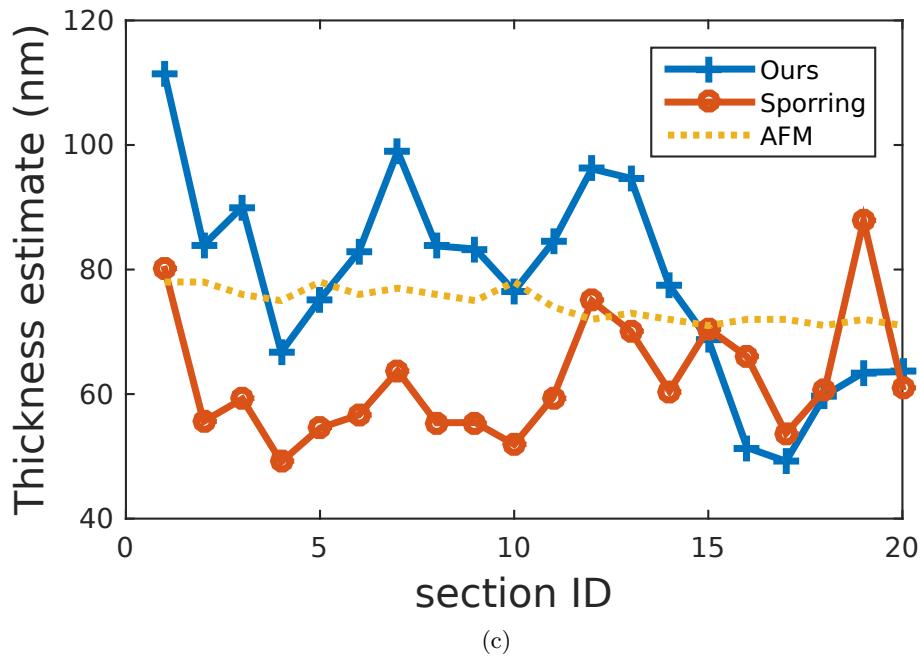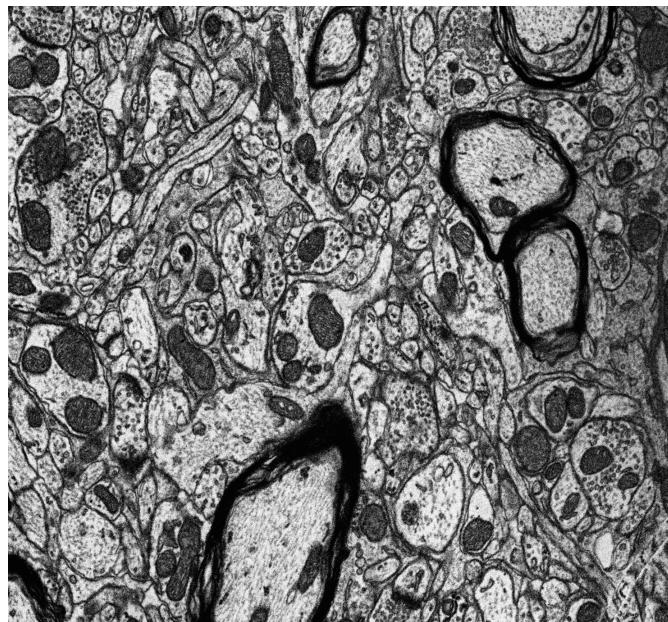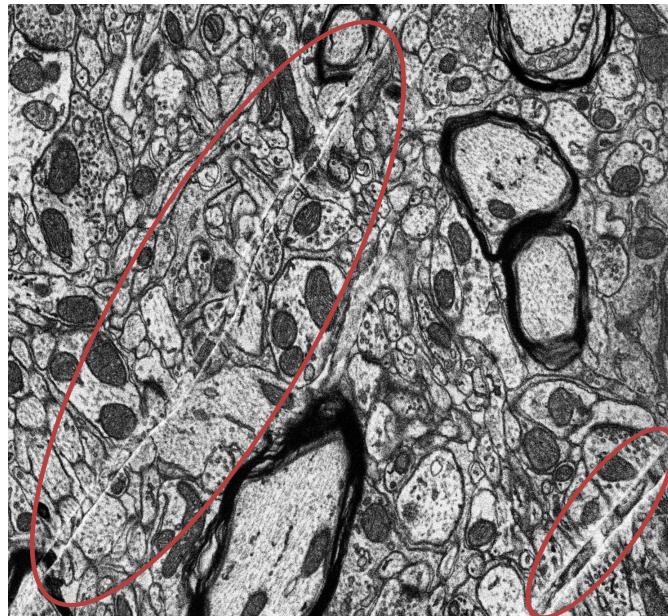
Figure 2.9: Validation of thickness estimates: Comparisons of thickness estimates using our method (GP), XY averaging method suggested by Sporring et. al. [74] and direct measurements of section thickness by atomic force microscopy. 20 serial sections sliced by an ultra-microtome and prepared for ssSEM was used. From those 20 serial sections, the AFM measurements were taken at the edge of each section as described in Section 2.9. Three different ssSEM stacks were obtained at different locations of the same 20 sections. Each of the above graphs correspond to one of those stacks.

(a)

(b)

Figure 2.10: Images corresponding to the peak ($sectionID = 14$) of the graph shown in Fig. 2.10b. This value is calculated based on the similarity between image (a) and image (b). Artifacts seen as white lines in image (b) (highlighted using red ellipses) have possibly contributed to increase the dissimilarity between the two images, thereby resulting in an overestimate of the thickness.
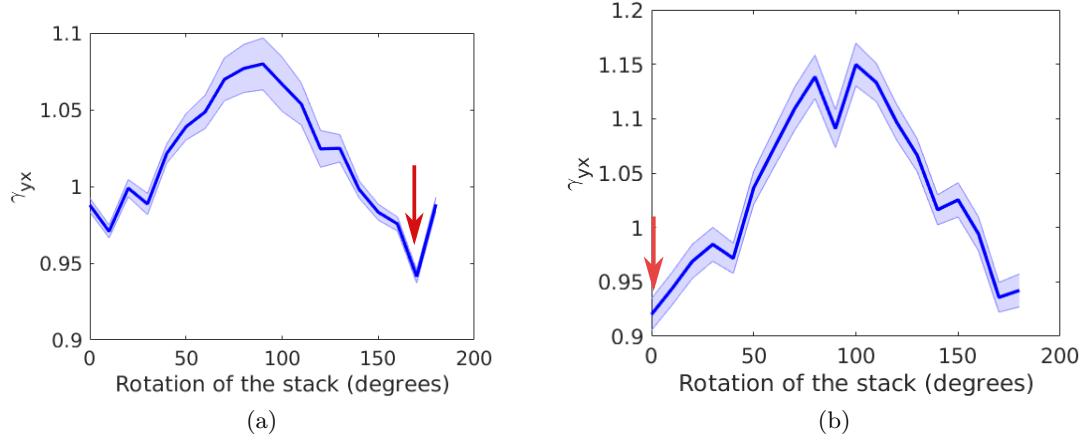
Figure 2.11: γ_{yx} estimated for two image stacks separately for different rotations about the Z axis. The lowest γ_{yx} (pointed out by arrow) corresponds to the stretching coefficient when the maximum linear compression is along Y axis. At this point, the direction along which the linear compression is minimal aligns with the X axis.
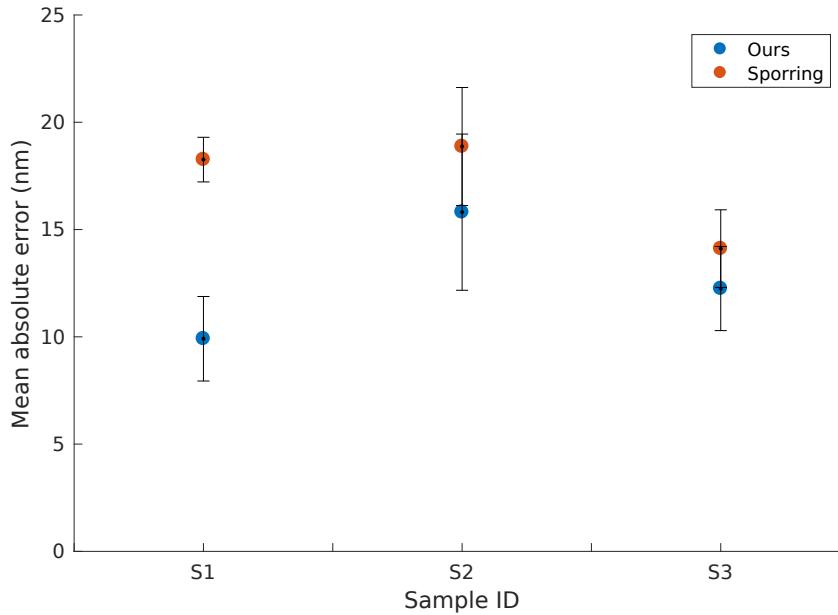
Figure 2.12: Absolute estimation errors using our method and [74] with respect to AFM measurements of section thickness for the three image stacks for which the estimates are plotted in Figure 2.9. Mean absolute error estimates for the three volumes using our method: $9.91\% \pm 1.97$, $15.81\% \pm 3.64$ and $12.25\% \pm 1.96$; using Sporring *et. al.* [74]: $18.26\% \pm 1.04$, $18.87\% \pm 2.75$ and $14.11\% \pm 1.81$.

2.11 Conclusion

We have presented a method to estimate both thickness and stretching in EM imagery, using image statistics alone. Our method is based on learning the distance between sections as a function of their dissimilarity. The stretching coefficient quantifies the cumulative effect of different sources of anisotropy along the xy plane including handling, storing, cutting, imaging or simply intrinsic anisotropy of specimen. Estimation of xy anisotropy is particularly useful to improve the accuracy of the thickness estimation due to its dependency on the pixel intensity variation along the xy plane. Moreover, it is suitable as a pre-processing step for any method that assume isotropy in the image statistics. As part of this work we have created a dataset of 20 ssSEM images along with thickness measurements directly obtained by AFM. We use this dataset to compare the performance of our thickness estimation method with other methods that use image statistics for indirect estimation of section thickness.

Thickness estimation methods based on image statistics alone are prone to inaccurate estimation of section thickness if sample anisotropy is not taken into account. Our anisotropy estimation method can be used to select the optimal rotation of the original image stack to train a regressor that is minimally affected by sample anisotropy so that it can be used to predict thickness. However this gives best results when sample inherent anisotropy is minimal. To circumvent this problem we suggest to use large enough images so that effects of any directed structured that might occur locally may even out given a larger scope.

Chapter 3

2D neuron slice segmentation using topological constraints

3.1 Introduction

The most common problems in automatic neuron segmentation are caused by local ambiguities at neuron membranes when the algorithms erroneously merge two adjacent neurons into one. Such ambiguities are a result of low SNR in some parts of the image either because of imperfections in the image acquisition process or lack of sharpness of neuron membranes. Human annotators are much better at resolving such problems due to their ability of considering a larger context when dealing with local ambiguities. Therefore, one way for automatic segmentation methods to reach human level segmentation accuracy would be to take into account a larger context considering shape cues and continuation of structures in the presence of gaps. We present a method to automatically annotate neuron slices (2D) on individual EM sections by accurately segmenting the neuron boundary that surrounds each neuron slice.

Generic classifiers can be used to classify pixels of EM images into classes such as neuron membrane, synapses and mitochondria. Pixelwise probability maps (Fig. 3.1b) thus obtained are commonly used as the main input in many neuron segmentation approaches [5, 22, 23, 46, 47, 58, 83].There has been a significant improvement in pixel level image classification due to the recent progress in the area of deep neural networks, in particular with convolutional neural networks (CNN) [13, 81]. The main drawback of CNNs is that they require a large training dataset which requires a significant amount of effort. When using CNNs, slight variations in the imaging parameters in datasets requires the CNNs to be trained separately for each dataset. Compared to CNNs, random forest classifiers (RFCs) require much less training data. However the quality of the probability maps generated by RFCs is usually less than those from CNNs. Our method produces significantly more accurate neuron segmentations using noisy probability maps from RFCs, compared to segmenting the same probability maps using standard techniques like the graph cut [7], thereby providing a way of generating reasonably accurate segmentations using a small amount of training labels.

The automatic segmentation method we propose uses topological constraints that reflect expected properties of an accurate segmentation of 2D neuron slices on an EM

(a)   (b)
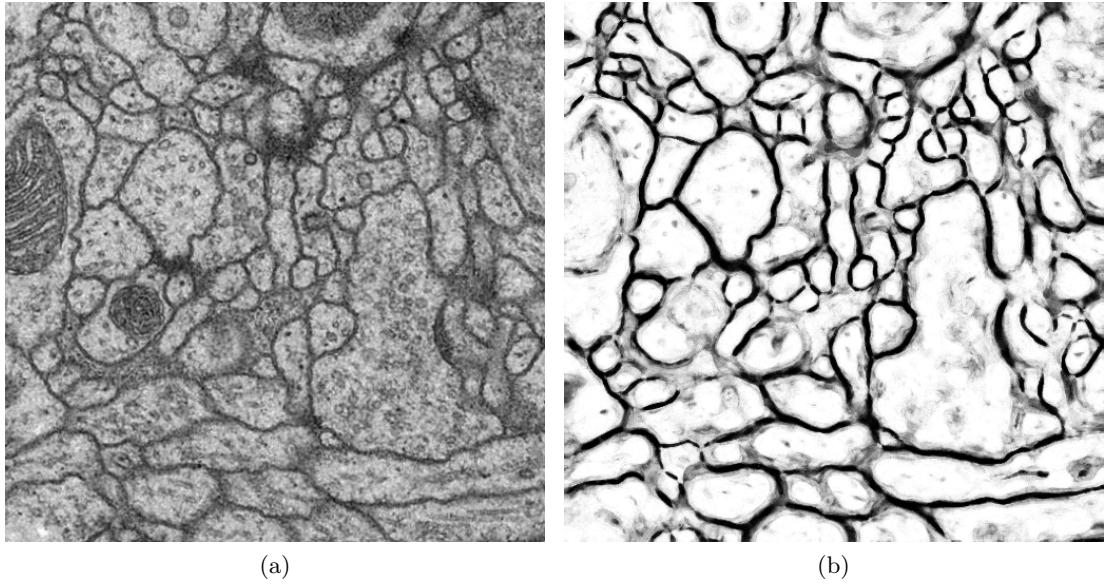
Figure 3.1: (a) ssTEM image of a section of the ventral nerve chord of drosophila larva. (b) Corresponding membrane probability map generated using a random forest classifier. Low pixel intensity (dark) corresponds to high membrane probability.

section. These constraints are defined on edges, nodes and faces of a planar graph which is derived from a membrane probability map. The segmentation problem is formulated as an integer linear program (ILP) that assigns an active or inactive state to each of the binary state variables (edges, nodes and faces of the graph) under these constraints. In addition to the constraints, we define an objective function that uses prior information about the state of these variables which is obtained by means of state of the art classifiers. Sec. 3.3 describes how we derive a graph of over-segmented object boundaries from a membrane probability map. Sec. 3.4 to Sec. 3.7 presents the formulation of our segmentation method as an ILP. Sec. 3.8 provides an evaluation of our approach with comparisons to other methods.

3.2   Related work

With the increase of automated and efficient EM image acquisition methods, there has been many advances in automated 3D neural circuit reconstruction methods in recent years [23, 83, 47]. Most of these methods commonly use pixelwise neuron membrane probabilities estimated using state of the art classifiers.

The method presented in this paper focuses on segmenting neurons on 2D images which is potentially useful in 3D reconstruction pipelines where the accuracy of 3D reconstructions depend on the quality of 2D candidate segmentations [23, 24]. Other methods that can be used to generate 2D segments include the graph cut and region merging approaches [58, 40, 5]. Among the region merging approaches [58, 5] work in both 2D and 3D.

Our approach differs from the above methods because we use local constraints that

focus on lowering topological errors which are usually caused by lower SNR, thereby addressing problems that are typically difficult for automatic methods.

3.3   Graphical representation of the segmentation problem

Our method essentially represents the neuron segmentation task as an edge labeling problem on a planar graph (Fig. 3.2), where the edges of the graph correspond to potential object boundaries which delineate cell interior and membrane segments. The edge labeling procedure results in assigning a state to each of those edges. The assigned state would either be *active* or *inactive*, such that each active edge will be considered a true object boundary i.e. the two faces of the graph on either side of that edge will be assigned to opposite classes one of which is *foreground* (cell interior) and the other *background* (membrane).

Following sub-sections describe how we represent a membrane probability map of an EM image as a planar graph (Fig. 3.3a) which is then used to formulate an ILP. The optimal solution of the ILP states which edges (and regions) are active as shown in Fig. 3.3b.

3.3.1   Probability maps

As the first step, we obtain a membrane probability map (Fig. 3.1b) of neuron membranes corresponding to the input EM image (Fig. 3.1a). Each pixel of such a probability map gives the probability of that pixel belonging to the class of neuron membrane. These probability maps can be generated using any generic classifier of choice (e.g. CNN, RFC).

3.3.2   Graph of over-segmented object boundaries

In the second step potential object boundaries are extracted from probability maps. We use a bank of oriented edge detection filters (Fig. 3.2a) to detect possible object boundaries along with their orientation. The output of such a filter bank applied to a membrane probability map quantifies how likely it is for an edge to be present at each pixel for each of the orientations defined in the edge filter bank. The filters have orientations ranging from 0° to 350° in steps of 10°. The maximal response from these filters at each pixel is illustrated in Fig. 3.2b.

A watershed transform is applied on the filter output to obtain a height map of the local maxima of the max response of the filter. Here the watershed transform implemented in MATLAB 2016b as formulated in [55] was used. This implementation results in locally connected components (or superpixels) of similar pixel intensities, as quantified by local intensity gradients. Each superpixel is assigned a unique identifier. A boundary having a width of one pixel is placed between any two superpixels with different identifiers. These boundaries are referred to as 'ridges'.

The ridges of the watershed transform corresponds to a graph of containing all possible edges detected by the oriented filter bank (Fig. 3.2c). Therefore, this graph can be considered as an over-segmentation of object boundaries that we would proceed to segment. To avoid having too many small watershed super-pixels (a.k.a. fragments), the

(a)   (b)   (c)
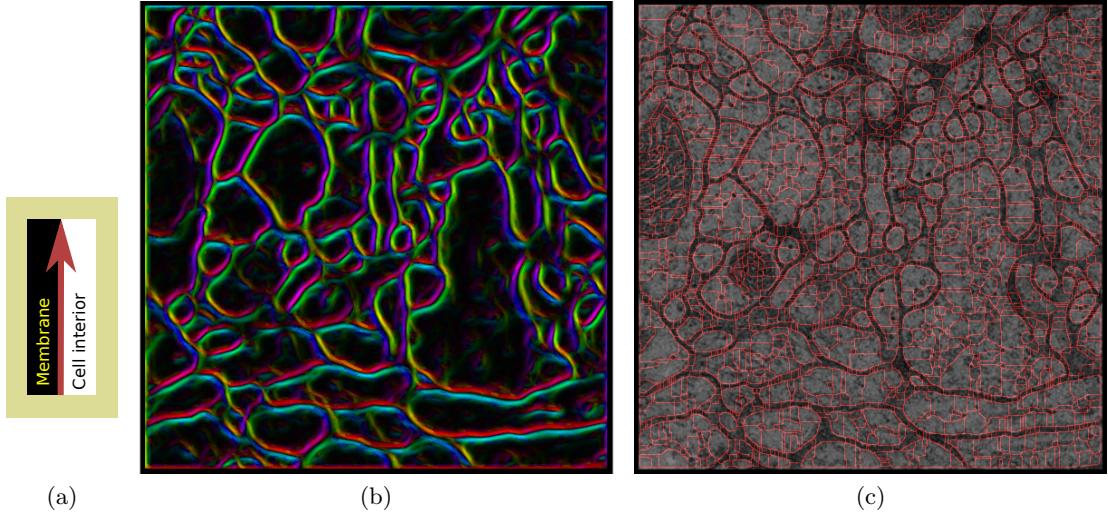
Figure 3.2: (a) Edge filter used for oriented edge filtering. 36 rotations (10° each) of this structure are convolved with the membrane probability maps to get the oriented edge filter response. Cell interior (foreground) is always on the right hand side of the directed edge. (b) Max response of oriented edge filtering of Fig. 3.1b. The intensity of each pixel is the max of the convolutions of all oriented edge filters. The hue $h \in [0,1]$ of each pixel corresponds to the orientation $\theta \in [0,360°]$ of the edge with the highest response. (c) Watershed transform of the max response of oriented edge filtering.

filter response is smoothened using a Gaussian filter with $\sigma = 1.4$ pixels, before applying the watershed transform.

3.4 Problem representation using binary state variables
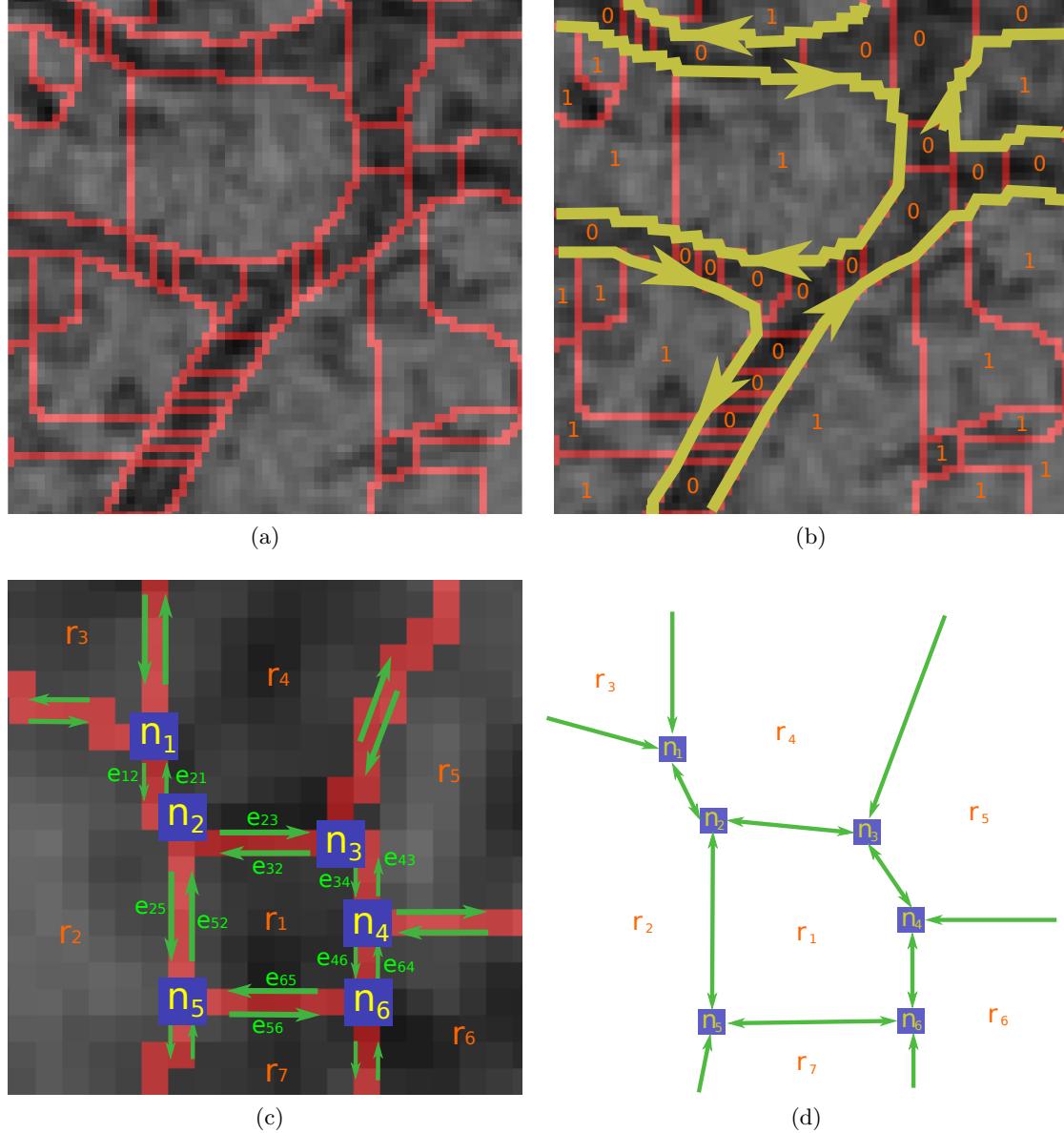
Figure 3.3: Illustration of the graph on which our optimization problem is defined: (a) Zoomed in version of Fig. 3.2c showing the extracted planar graph representing possible object boundaries (red) that delineate cell interior and neuron membranes. This graph is overlaid on the EM image from which the edges are derived. (b) Expected outcome of our segmentation task. Each region which is part of cell interior is assigned 1 and membrane is assigned 0. Each active (directed) edge separating cell interior and membrane is highlighted in yellow. All inactive edges remain red. (c) Binary state variable types used in Sec. 3.4: *edges (e), nodes (n)* and *regions (r)* to define an integer linear program are annotated on a zoomed in version of (a). Note that each object boundary in Fig. (a) (shown in red) between any two nodes n_i and n_j is denoted by a pair of edges having opposite directions (green). (d) Sketch of the same graph shown in (c).

Using the graphical representation in Fig. 3.3a we formulate an optimization task that results in the desired labeling of edges and regions where the regions which are part of cell interior are assigned to foreground and the regions which are part of membrane are assigned to background as illustrated in Fig. 3.3b. This section describes how we formulate an ILP using the graph of over-segmented object boundaries such that the optimal solution corresponds to an accurate segmentation of neuron boundaries.

We define three types of binary state variables corresponding to the edges, nodes and faces of that graph as shown in Fig. 3.3d. The *active state* of a binary state variable means that it has the value one.

1. **Edge state binary variables** Each edge shown in the graph (Fig. 3.3a) between two nodes n_i and n_j colored in red has three corresponding state variables: (a) $e_{ij} = 1$: directed edge exists from node n_i to n_j, (b) $e_{ji} = 1$: directed edge exists in the opposite direction, (c) $e_{ij}^0 = 1 \iff e_{ji}^0 = 1$: lack of an edge between nodes n_i and n_j. If the edge between the nodes id j is part of the object boundary, either e_{ij} or e_{ji} is set to be active by the ILP (Fig. 3.3b).

2. **Region state binary variables** Each face (region) m of the graph has two corresponding state variables r_m and r_m^0. $r_m = 1$ when region m is part of foreground in the segmentation output. r_m^0 is active when region m is part of background.

3. **Node state binary variables** In an accurate segmentation output, the angles between two edges (neuron boundaries) tend to be smooth. Therefore, sharp angles occurring between two active edges are penalized in the ILP objective. Assigning such a penalty to the activation of a pair of edges gives rise to a quadratic program. In order to keep the problem formulation linear we define *node state variables* which correspond to pairs of directed edges. At any node of the graph (Fig. 3.3a) there are at least 3 edges (6 directed edges as in Fig. 3.3d) out of which exactly two or zero has to be active as described later in Sec. 3.6. A set of node state variables n_i^c are defined for each node. When a particular node state n_i^c is active, it specifies which pair of directed edges are active out of all edges connected to this node.

   $n_i^{c=0}$ is the state where the node is not active i.e. not part of the segmentation and therefore none of the edges connected to it are active. Fig. 3.4 illustrates the set of node state variables for a typical node of the graph with 3 neighbors.

3.5   Integer Linear Program

Linear Programming (LP) is a technique for optimizing a set of state variables that can take a continuous set of values within the given constraints for the problem by optimizing an objective function which is a linear combination of the state variables. Integer Linear Programming is a special case of Linear Programming where the state variables can only take integer values.

The cannonical form of an ILP is as follows.

$$\max_{x} c^T x, \tag{3.1}$$

subject to:

$$Ax < b, \tag{3.2}$$

and

$$x \in \mathbb{Z}. \tag{3.3}$$

Here c is the known coefficient vector for the state variables x in the objective function. A is the known coefficient matrix of the constraints and b is a set of known constants. Depending on the structure of the problem, ILPs with up to hundreds of thousands of state variables can be efficiently solved with current state of the art optimization tools [29]. The structure of the problem relates to the set of constraints and the objective function of the optimization task. In our case optimal solutions can be found in the matter of seconds suggesting that the problem structure has a favorable form for the solution to be found efficiently.

3.6 ILP constraints

The constraints of an ILP are a set of bounds for different linear combinations of the state variables that define the solution space of the entire set of state variables. We define a set of hard constraints that reflect general properties expected in an accurate segmentation output. Following is a description of these linear constraints used in our ILP.

3.6.1 Low-level constraints

We define an activation constraint for each variable type as follows.

1. **Directed edge activation** Between two adjacent nodes n_i and n_j, we allow either the outgoing edge e_{ij} (w.r.t n_i) to be active or the incoming edge e_{ji} to be active or none of them to be active ($e_{ij}^0 = 1 \iff e_{ji}^0 = 1$). Therefore, each edge of the graph in Fig. 3.3a has three corresponding binary edge states, out of which exactly one has to be set to 1. The mathematical formulation of this constraint is as follows:

$$\forall e_{ij} \in \mathcal{E}, \ e_{ij} + e_{ji} + e_{ij}^0 = 1, \tag{3.4}$$

where $\mathcal{E}$ is the set of all edges in the graph.

2. **Node activation** Similarly, for each node n_i we enforce that one of all possible nodes states ($c \in \{0, 1, ..., p\}$) must be active.

$$\forall n_i \in \mathcal{N}, \ \sum_{c=0}^{p} n_i^c = 1, \tag{3.5}$$

where $\mathcal{N}$ is the set of all nodes and p is the number of binary node states of node n_i as illustrated in Fig. 3.4.

3. **Region activation** Each face of the graph (region) is assigned two binary indicator variables: r_m^1 and r_m^0. When region m is assigned to foreground, the
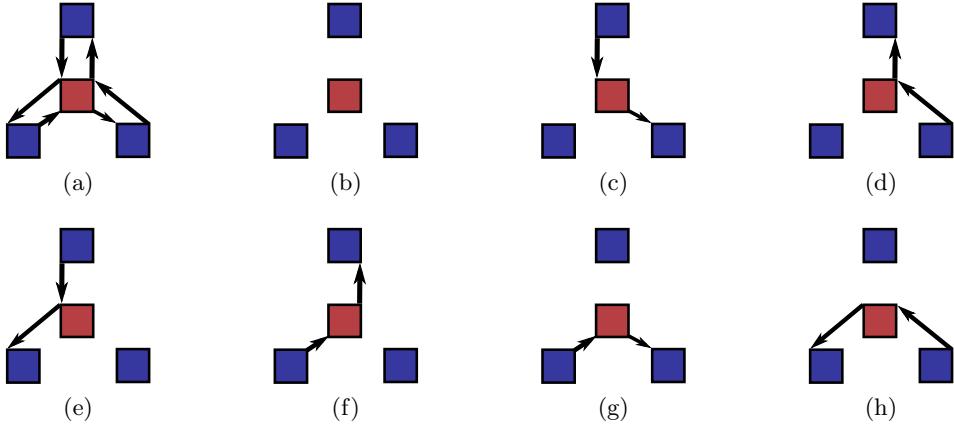
Figure 3.4: *Node states*: (a) depicts a typical node (in red) in a graph similar to Fig. 3.3d that is connected to each of its three neighboring nodes (blue) via a pair of directed edges. Fig.(b) to (h) are all possible p node states for the node in red. Each of these states are represented by a binary variable in the ILP formulation. (b) shows the inactive node state where it is not connected to any of its neighbors. (c) to (h) show all possible active states this node is allowed to have according to the given constraints. In each of these active node states there are exactly two active edges attached to it, where one is incoming and the other out going. Therefore, $p = 2 \cdot \binom{L}{2} + 1$ (where L is the number of directly connected nodes to the current node).

corresponding region binary state variables should have the values: $r_m^1 = 1$ and $r_m^0 = 0$. Therefore, region state variables are activated according to the following constraint:

$$\forall m \in \mathcal{R}, \ r_m^1 + r_m^0 = 1, \tag{3.6}$$

where $\mathcal{R}$ is the set of all regions.

3.6.2 Topological constraints

Co-activation of different combinations of state variables results in different topologies in the output segmentation. If not properly constrained, the resulting topology could deviate from what's expected in an accurate segmentation. Therefore, we define the following constraints in order to ensure that resulting segmentation have an acceptable topology.

1. **Closed loop of edges** In the segmentation output each foreground segment must be enclosed by a closed loop of edges. This also means there must not be any dangling edges activated, that do not have exactly two other edges connected to either end of it (Fig. 3.5a). We enforce this topological property as a constraint as follows:

   - When a node is active exactly two edges connected to it have to be active as well.
   - When the node n_i is inactive ($n_i^0 = 1$), none of the edges attached to it can be active.

(a)   (b)
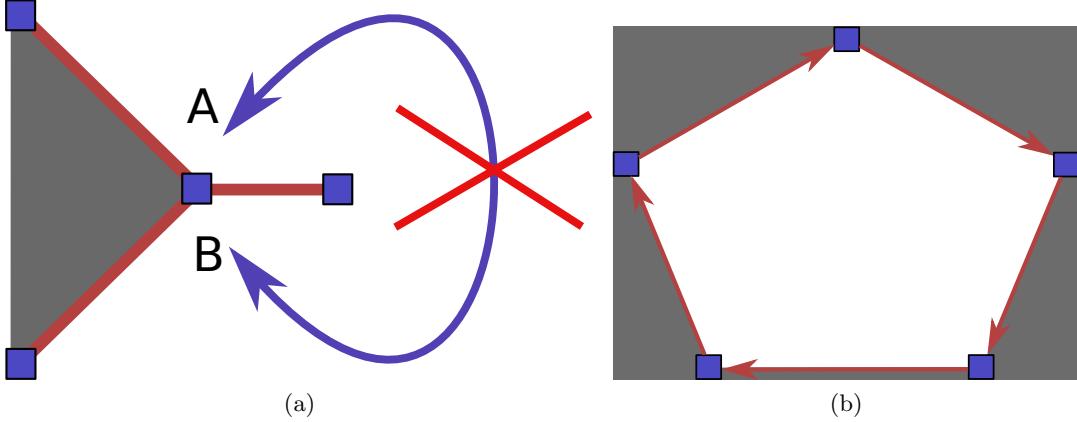
Figure 3.5: (a) A dangling edge would mean that the regions A and B on either side of the edge are of the same type. This is a topological error because an edge has to have two opposite types of regions on its sides (foreground and background). We prevent this using the constraint *closed loop of edges* (Eq. 3.7). (b) Because of our definition of a directed edge, any foreground segment has to be bounded by a closed clockwise loop of active edges. We enforce this topological property using constraint *clockwise edge loop around foreground* (Eq. 3.9).

This results in the following mathematical formulation:

$$\forall n_i \in \mathcal{N}, \ 2n_i^0 + \sum_{e_{ij} \in E_{n_i}} e_{ij} = 2, \tag{3.7}$$

where E_{n_i} is the set of edges attached to node n_i. Any neighbor of n_i connected via the edges e_{ij} and e_{ji} is denoted by n_j.

2. **Activate corresponding edges of a node state** For each active state of node n_i, there is a unique pair of directed edges that have to be activated. While constraint (3.7) ensures there are exactly two active edges for an active node, this constraint (Eq. (3.8)) ensures the matching between the predefined node states and the corresponding active edge pairs. This matching is important because in the objective function (Sec. 3.7) we use a precalculated set of costs for each node state based on the angle between its active pair of edges.

Consider a typical node n_i with (at least) a pair of neighboring nodes n_j and n_k. Let node state n_i^c have the edges e_{ij} (outgoing edge) and e_{ki} (incoming edge) activated.

We formalize this constraint as follows:

$$\forall n_i \in \mathcal{N}, \forall c \in \mathcal{C}_{n_i}, \ 0 \leq -2n_i^c + e_{ij} + e_{ki} \leq 1, \tag{3.8}$$

where $\mathcal{N}$ is the set of all nodes and $\mathcal{C}_{n_i}$ is the set of all node states of node n_i.

3. **Clockwise edge loop around foreground** As mentioned previously, topological constraint (i) given by Eq. (3.7) enforces a closed cycle of edges around every foreground segment. However according to the definition of a directed edge (Fig. 3.2a)
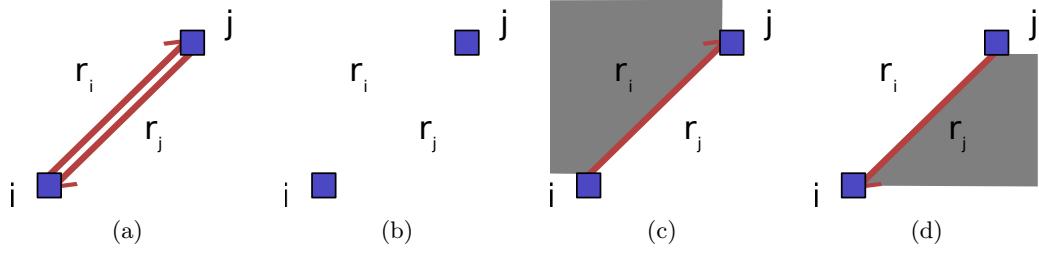
Figure 3.6: (a) Sketch of two nodes i and j with two directed edges e_{ij} and e_{ji} between them. Region to the left of e_{ij} is r_i where as the region to its right is r_j. (b) Clockwise edge loop constraint given in Eq. (3.7) enforces both r_i and r_j to be of the same state if the edge in between the regions is inactive. (c) and (d) If one of the directed edges is active, the same constraint enforces that the region on the left of the directed edge is of type membrane.

foreground must occur to the right of an active edge. Therefore, these closed cycle of edges should have a clockwise sense (Fig. 3.5b).

This topological constraint is formulated in the ILP as follows:

$$\forall e_{ij} \in \mathcal{E}, \; -e_{ij} + e_{ji} + r_j^1 - r_i^1 = 0. \tag{3.9}$$

4. **Membranes as closed loops** We do not expect to see gaps in the segmented neuron membranes. In other words, as shown in Fig. 3.3b, the segments which are assigned to be membrane (background) are contiguous. Any region which is assigned to background must have at least two other adjacent regions connected to it that are also assigned to background. This property is formalized as a constraint using the following equation:

$$\forall r_i \in \mathcal{R}, \; r_i^1 - r_i^0 + \sum_{e_j \in E_{r_i}} e_j^0 \geq 1. \tag{3.10}$$

Here, e_j is an element of the set E_{r_i} of all edges bounding region r_i and $e_j^0 = 1$ when edge e_j is not active. In Eq. (3.10) making sure that at least 2 of the edges are inactive when the region is turned off (i.e. assigned to background), we enforce that region to be part of a contiguous membrane segment.

We observe that this constraint helps the system to fill gaps observed in membrane probability maps (Fig. 3.1a). This is useful since neuron membranes sometimes gets blurred out and leads to merging of two adjacent neuron sections via this blurry region due to that region being misclassified as being part of cell interior.

3.7   ILP objective function

The purpose of the objective function of our ILP is to minimize the deviation of the optimal solution from prior information derived from the data. We precompute coefficients for each binary state variable using the input data, using classifiers that are trained on
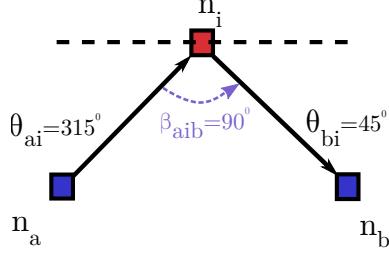
Figure 3.7: Illustration of node angle β_{aib} used in Equation 3.14. θ values are obtained from *oriented edge filters* as described in Sec. 3.3.2. The horizontal dotted line is the reference with respect to which the angles are defined for the oriented edge filters.

available ground truth. The objective function of the ILP is:

$$\{e^*, n^*, r^*\} = \underset{e \subset \mathcal{E}, n \subset \mathcal{N}, r \subset \mathcal{R}}{\operatorname{argmin}} \quad C_e + C_n + C_r. \tag{3.11}$$

Here, the terms C_e, C_n and C_r represent the weighted cost terms for the activation of the state variables corresponding to edges, nodes and regions in the model respectively. The objective term C_e for edge state variables is defined as:

$$C_e = \sum_{e_i \in \mathcal{E}} w_e^{on} \cdot u_{e_{ij}} (e_{ij} + e_{ji}) + w_e^{off} \cdot (1 - u_{e_{ij}}) \cdot e_{ij}^0, \tag{3.12}$$

where $w_{e_{on}}$ and $w_{e_{off}}$ are the linear weights controlling the relative reward (or penalization) of turning an edge on (or off). $u_{e_{ij}}$ is the *a priori* probability for edge e_{ij} to be active i.e. be part of an object boundary. This probability can be learned using any general classifier of choice. Fig. 3.8a illustrates *a priori* edge activation probabilities obtained using a random forest classifier.

The cost term C_n corresponding to all node state variables $n_i \in \mathcal{N}$ is:

$$C_n = \sum_{n_i \in \mathcal{N}} \{w_n^{off} \cdot u_{n_i^0} \cdot n_i^0 + w_n^{on} \sum_{c \in \mathcal{C}_i} u_{n_i^c} \cdot n_i^c\}, \tag{3.13}$$

where $\mathcal{C}_i$ is the set of all node states of node n_i.

In Eq. (3.13) we have set the score $u_{n_i^0}$ for any inactive node configuration to be equal to 1 so that in the ILP objective its contribution is solely determined by the weight w_n^{off}. The score $u_{n_i^c}$ assigned to each active node configuration c of node n_i is calculated as

$$u_{n_i^c} = f(\beta_{aib}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta - \pi)^2}{2\sigma^2}} \tag{3.14}$$

where β_{aib} is the angle between the two active edges corresponding to the active node configuration j of node n_i (Fig. 3.7). This score is a smoothness term which is maximal for a node configuration where the incoming and outgoing edges are at an angle 180° and drops to zero when it deviates from 180°. The variation of the smoothness is modeled using a Gaussian function with mean 180° and a predefined spread σ ($\approx 45°$) as given in equation 3.14.

The objective term for all regions C_r is:

<div align="center">(a)   (b)</div>
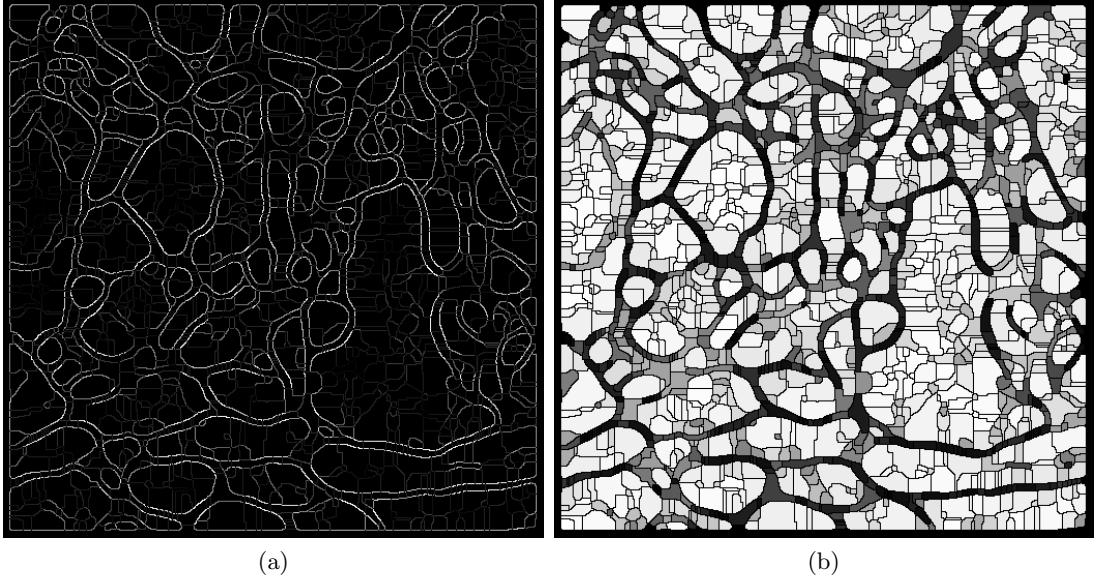
Figure 3.8: Visualization of *a priori* probabilities associated with *edge state variables* and *region state variables* of an EM image, which are used in the ILP objective (Equation. 3.11). (a) Each pixel belonging to an edge indicates the *a priori* probability of that edge being at the boundary between neuron membrane and cell interior. These probabilities are generated using a random forest classifier. White indicates a high probability. (b) Each region is assigned an *a priori* probability that it belongs to cell interior. This probability is obtained from the membrane probability map by averaging the pixelwise probabilities over each region.

$$C_r = \sum_{r_i \in \mathcal{R}} w_{r_i}^{on} \cdot u_{r_i} \cdot r_i + \sum_{r_i^0 \in \mathcal{R}} w_{r_i}^{off} \cdot u_{r_i^0} \cdot r_i^0 \tag{3.15}$$

where u_{r_i} is the *a priori* probability of the region being part of cell interior. This value is obtained by averaging the membrane probability of all the pixels in a given region (Fig. 3.8b).

3.7.1   Structured learning of cost function parameters

The linear cost function (Equation 3.11) used in the ILP has six linear weights w_e^{on}, w_e^{off}, $w_n^{on}, w_n^{off}, w_r^{on}$ and w_r^{off} as described in expansions of the cost terms in equations 3.12, 3.13 and 3.15. We learn optimal values for these weights using *structured learning*.

The optimal value of the ILP given by Equation 3.11 can be seen as the state $\mathbf{y}^*$ that corresponds to the minimal energy of $E(\mathbf{x}, \mathbf{y}; \mathbf{w})$ where $\mathbf{x}$ is the observations vector. Here, $\mathbf{y}$ is a vector containing all the state variables (edges, nodes and regions) based on which the optimization task is formulated as described earlier. The weight vector $\mathbf{w}$ is used to linearly weight the real valued features $\phi(\mathbf{x}, \mathbf{y})$ (Equation 3.17). In our case these real valued features correspond to the terms in the objective function that are to be linearly weighted.

$$f(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y} \in \mathcal{Y}} E(\mathbf{x}, \mathbf{y}; \mathbf{w}) \tag{3.16}$$

$$E(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \langle \mathbf{w}, \phi(\mathbf{x}, \mathbf{y}) \rangle \tag{3.17}$$

Under the structured learning frame work we find the optimal set of weights $\mathbf{w}^*$ that minimises the empirical risk $J(\mathbf{w}, \mathbf{x}', \mathbf{y}')$ for a given training data set with labels $\mathbf{y}' \in \mathcal{Y}$ and observations $\mathbf{x}' \in \mathcal{X}$.

$$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} \; J(\mathbf{w}, \mathbf{x}', \mathbf{y}') \tag{3.18}$$

$$J(\mathbf{w}, \mathbf{x}', \mathbf{y}') = \Omega(\mathbf{w}) + L(\mathbf{w}, \mathbf{x}', \mathbf{y}') \tag{3.19}$$

$L(\mathbf{w}, \mathbf{x}', \mathbf{y}')$ is the loss function defined on the training data set which relates to *hinge loss* in *maximum margin training*. $\Omega(\mathbf{w})$ is a convex regularization term defined on the weight vector $\mathbf{w}$.

$$L(\mathbf{w}, \mathbf{x}', \mathbf{y}') = \max y \in \mathcal{Y} \tau(\mathbf{y}', \mathbf{y}) \delta E(\mathbf{y}', \mathbf{y}) + \Delta(\mathbf{y}', \mathbf{y}) \tag{3.20}$$

where $\Delta(\mathbf{y}', \mathbf{y})$ is an application specific cost, for which we used Hamming distance between $\mathbf{y}'$ and $\mathbf{y}$ and $\delta E(\mathbf{y}', \mathbf{y})$ is defined as follows:

$$\delta E(\mathbf{y}', \mathbf{y}) = E(\mathbf{x}', \mathbf{y}') - E(\mathbf{x}', \mathbf{y}) \tag{3.21}$$

For learning the weight vector $\mathbf{w}$ we used an implementation of structured risk minimization publicly available.[1]

3.8   Results

Evaluation criteria

The quality of neuron segmentation methods is usually calculated by comparing the segmentation results of those methods against a groundtruth labeling of neurons. We used Rand Index (RI) [63], Variation of Information (VoI) [54] and Tolerant Edit Distance (TED) [25] to evaluate our approach. It has to be noted that in the case of neuron segmentation, small shifts of the boundary of neurons can be ignored as long as such a shift will not result in a merge or split error. Both RI and VoI are known to be affected by boundary shifts of the segmentations [39, 25]. Therefore TED is more suitable for evaluating segmentation results in terms of their topological accuracy.

Topological error quantification using TED

TED quantifies topological errors of a given neuron segmentation against a given ground truth labeling of neurons. The topological errors evaluated are: false splits (one neuron segment assigned with two neuron labels), false merges (two different neurons assigned with the same neuron label), false positives (foreground segment identified where there should be just background), false negatives (background segment identified where there should be foreground).

---

[1]https://github.com/funkey/sbmrm
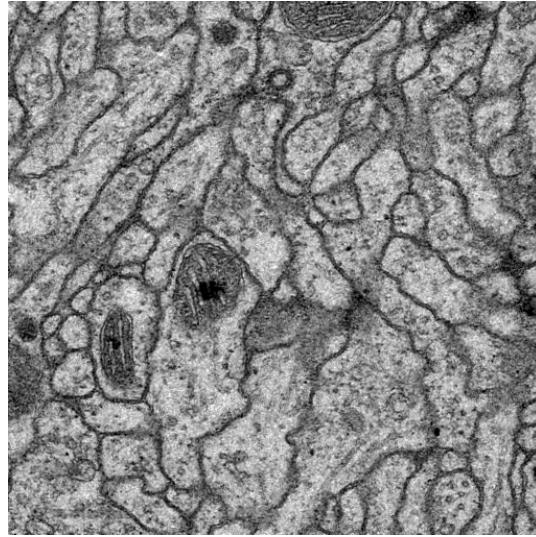
Figure 3.9: ssTEM image corresponding to the results shown in Figure 3.10. Image taken from ISBI 2012 [9] dataset.

Given two labelings y and g where g can be the ground truth segmentation, TED quantifies the topological error as the minimum number of possible false splits, false merges, false positives and false negatives while allowing the boundaries of the input segmentation y to shift by a small predefined threshold θ which is referred to as the tolerance. This tolerance criterion is used to produce different relabelings of y which are different versions of y that have boundary shifts smaller than the predefined tolerance.

The general formulation of TED between the labelings g and y is therefore:

$$TED(g,y) = \min_{y' \in Y(y)} splits(g,y') + merges(g,y'), \qquad (3.22)$$

where $Y(y)$ is the set of all relabelings of y within the allowed tolerance. These relabelings are found by obtaining superpixels (or supervoxels in 3D) whose labels can be changed. Such superpixels can be identified by intersecting the groundtruth labeling g and the segmentation output to be evaluated y. To find the minumum topological error over all possible tolerant relabelings, an integer liner program is used to solve Equation 3.22 as detailed in [25].

Experiments

We evaluate our method using a publicly available dataset[9] of neural tissue of Drosophila larva, along with manually annotated neuron membrane labels. This dataset contains 30 sections of serial section Transmission Electron Microscopy (ssTEM) images of size 512×512 pixels. Each pixel is of dimensions $4\,\text{nm} \times 4\,\text{nm}$. Each section is approximately $50\,\text{nm}$ thick. We set aside images 21 to 30 for evaluations.

(a)

(b)

(c)

(d)
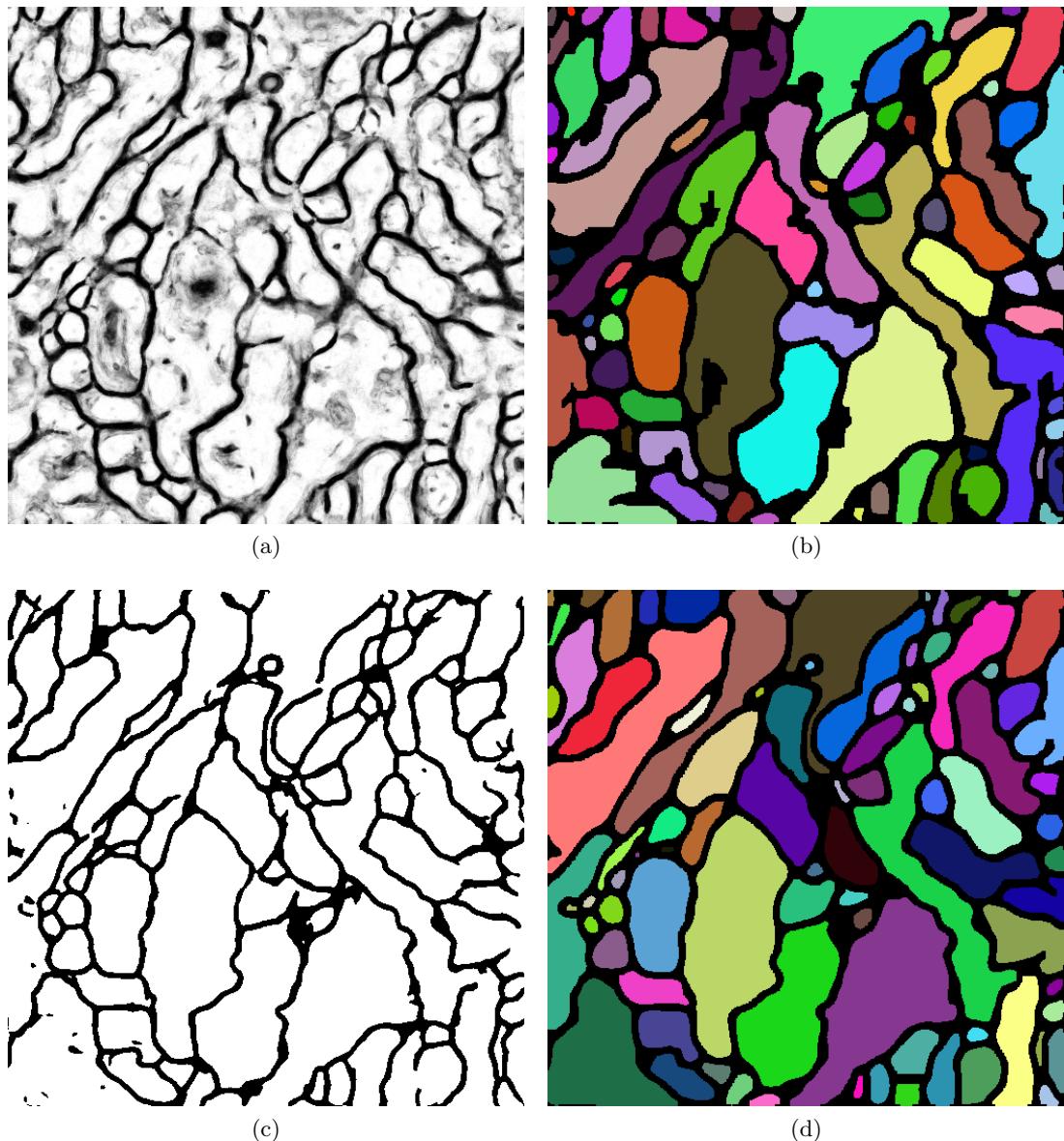
Figure 3.10: Segmentation outputs produced by our method using RFC probability map and CNN probability map for a single test EM image of size 512×512 pixels for the raw EM image shown in Figure 3.9. (a) Probability map obtained using RFC for a test EM image. (b) Segmentation output using our method for RFC probability map. (c) Probability map obtained using a CNN. (d) Segmentation output using our method for CNN probability map.

(a)

(b)

(c)

(d)
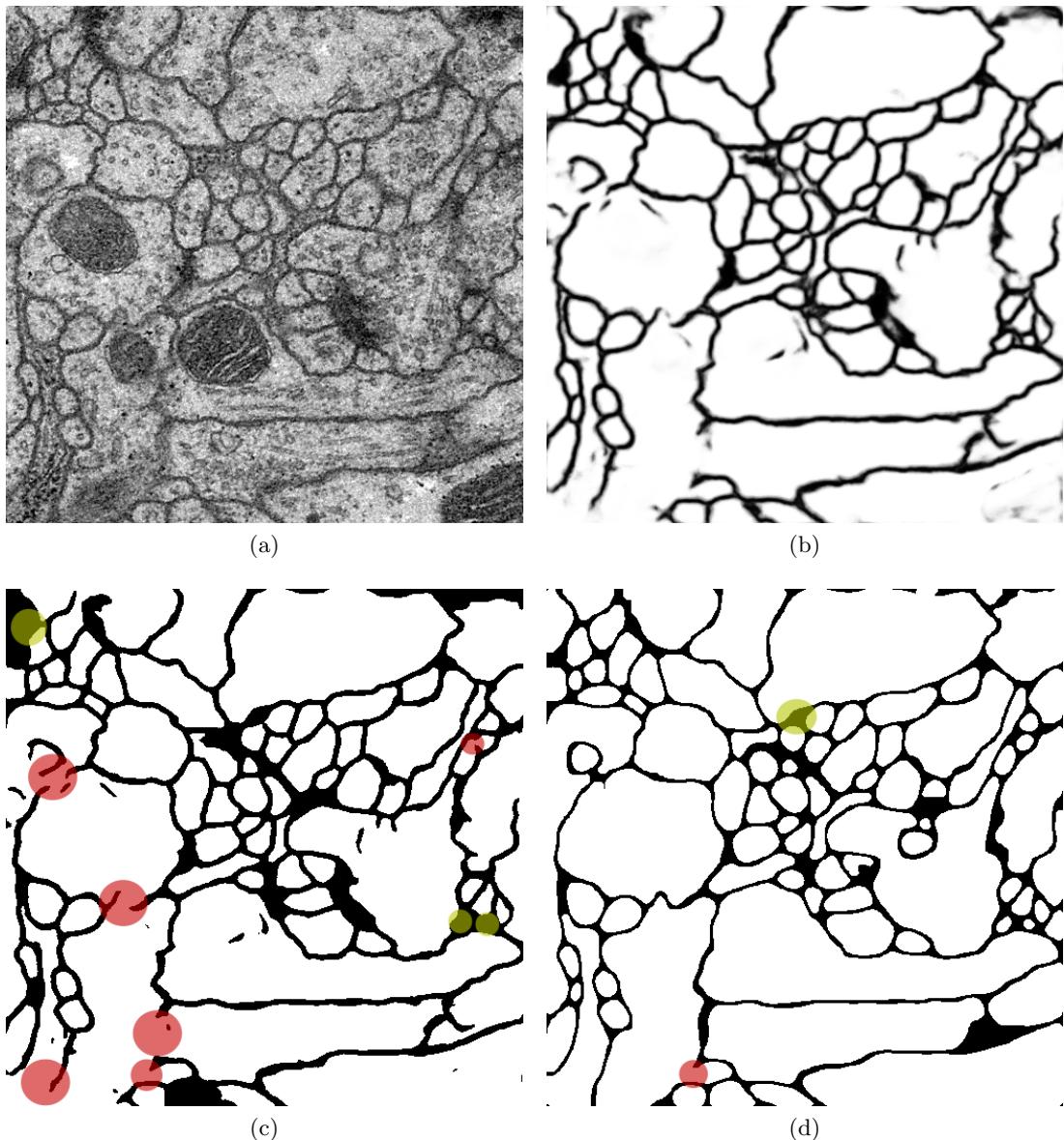
Figure 3.11: Comparison of topological errors in segmentation outputs produced by our method vs GALA [58] (region merging). (a) Raw EM image from ISBI 2012 [1] public dataset. (b) Membrane probability map produced by CNN. (c) Segmentation output produced by GALA for the same image. Foreground (cell interior) is shown in white where as background (membrane) is shown in black. Merge errors are shown in red. Missed segments (also known as false negatives) are pointed out in yellow. (d) Segmentation output using our method. Errors are highlighted with the same color code.
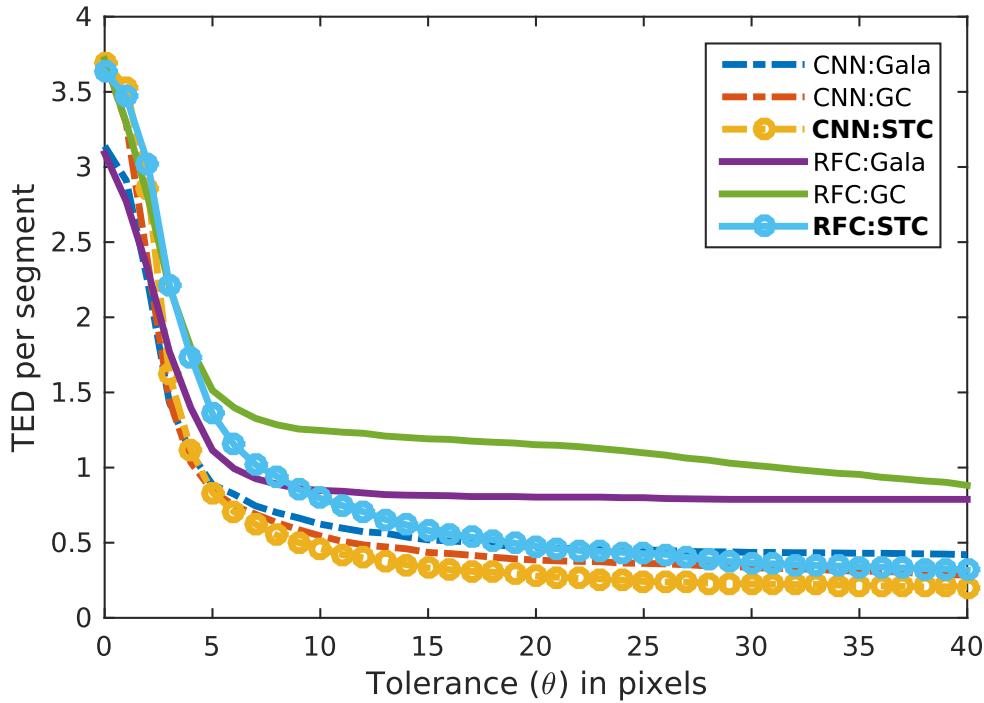
Figure 3.12: Total topological error quantified using TED plotted against the boundary tolerance allowed in pixels. which is the sum of false positives, false negatives, false splits and false merges. normalized by the number of segments in ground truth. Segmentation using our approach with CNN probability maps as inputs (CNN:STC) shows the best accuracy.

Probability maps

Using random forest classifiers

We used the random forest classifier provided by *Ilastik* segmentation toolkit [73]. The classifier was trained interactively by placing a few brush strokes to mark neuron membrane and cell interior separately on the first three images of the dataset. The reason to use such a small dataset for training is partly because the training accuracy doesn't significantly increase by adding more training labels for this classification task using RFCs. Furthermore, it is to point out that when there is not much training data available, RFCs are a good choice for pixel classification that still produces good results with our constraint-based segmentation method. Training of the RFC took less than five minutes. After training, processing each image to predict membrane probabilities took around 10 seconds.

Using convolutional neural networks

Another set of probability maps were generated using the CNN implementation provided by [81], using the first 20 images of the same dataset for training. This CNN implementation based on the Caffe CNN library [43] is an efficient implementation of Strided Kernel network architecture (SK-net) first suggested in [51]. The efficiency gain
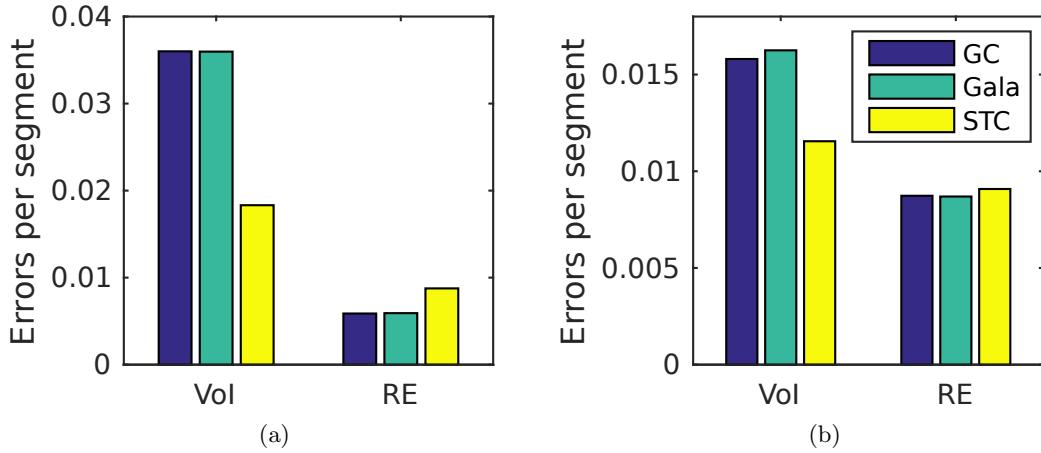
Figure 3.13: Comparison of segmentation errors using Variation of Information (VoI) and Rand Error ($RE = 1 - RI$). Methods compared: Graphcut (GC) [7], Graph-based active learning of agglomeration (GALA) [58] and our method (STC). (a) Using RFC probability maps as inputs. (b) Using CNN probability maps as inputs.

in SK-net as opposed to commonly used sliding window (SW) architecture comes from the fact that SK-net allows to predict serveral pixel labels at once instead of one pixel at a time. This is achieved by considering the overlapping context for each pixel that can be used to reduce redundant computations. Furthermore, the implementation by [81] takes advantage of parallelization provided by modern graphical processing units (GPU).

SK-net architecture has 9 hidden layers altogether. Three of the layers are convolutional layers combined with ReLU (rectified linear units). The other layers are three maxpooling layers and three inner product layers. Training of 20 images of size 512x512 used for comparisons took around 12 hours. Prediction using the trained network took less than one second per image of the same size.

Comparisons

We compare the performance of our method with the graphcut (GC) [7] and Graph-based active learning of agglomeration (GALA) [58] using probability maps generated as mentioned above. We have a lower VoI than both GC and GALA. Furthermore, our approach results in a lower cumulative topological error than GC and GALA (Fig. 3.12).
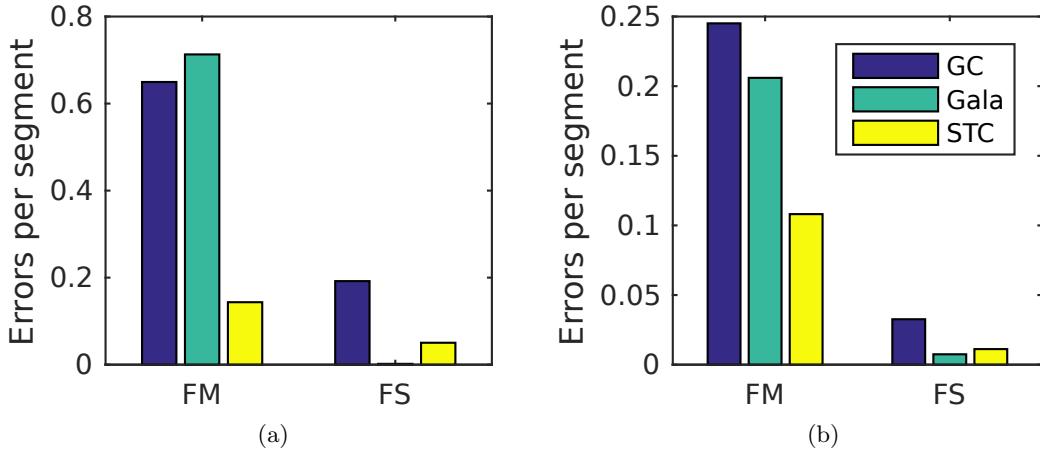
Figure 3.14: Breakdown of topological segmentation errors quantified using Tolerant Edit Distance (TED) [25] comparing Graphcut (GC) [7], Graph-based active learning of agglomeration (GALA) [58] and our method (STC). (a) False merges and false splits using RFC inputs. (b) False merges and false splits using CNN inputs.

3.9   Conclusion

We have proposed a method to segment neuron membranes in EM images which results in fewer topological errors. Our method takes pixelwise membrane probability maps as inputs which are then used to represent the segmentation task as an edge labeling problem on a graph. The edge labeling is solved by formulating an ILP using topological constraints to improve segmentation accuracy.

We have shown that our method can be used to produce 2D neuron segmentations using probability maps from both CNNs and RFCs with less topological errors than other segmentation methods. State of the art classifiers based on deep learning such as CNNs that produce high quality probability maps for EM datasets need a lot of training labels for each dataset, which is a major bottleneck for automated segmentation methods. On the other hand, random forest classifiers require much less training data with the drawback of resulting in probability maps of lower quality.

Because of the generic nature of the topological constraints used in our approach, we note that this method can be potentially used in any image segmentation pipeline that have similar properties.

Part II

Investigation of effects due to sensory exposure on zebra finch HVC using EM images

Chapter 4

Introduction to vocal learning in zebra finch

The ability to learn allows an organism or an artificial system to acquire new behaviours and adapt to its environment. It can be viewed as a means of expanding beyond what's biologically innate to an organism. The brain's capacity to learn is a highly interesting area in neuroscience research. Such research in neural basis of learning include studies of learning speech and other vocalizations. Apart from humans, no other primate species is known for learning complex vocalizations. Whales, dolphins and some types of bats are the rest of the species of mammals that exhibit vocal learning. However there are many species of songbirds that show this behaviour that provides a model to study the learning of vocalizations.

It has to be noted that speech learning and song learning have many differences. However, the common factor between the two is the sensorimotor learning of an elaborate vocalization. Bird songs are not known to have the property of a spoken language which is a medium for communicating complex messages. This is one obvious limitation of the analogy of song learning to human speech learning when songbird models are used as a potentially analogous system to study speech learning in humans.

This chapter gives an overview of the song system of zebra finch (*Taeniopygia guttata*) in order to provide a basis for the analysis carried out in the following chapter where we look into quantitative changes in a particular type of ultrastructure in a brain region in the song learning pathway of zebra finch.

4.1   Behavioural basis of song learning

The singing behaviour in zebra finch is only present with the males. The female birds only vocalize calls which is essentially a short announcement saying 'I'm hear'. Calling is common to both male and female zebra finches. The male juvenile birds learn to imitate a song from an adult male referred to as a tutor. At the earliest stage the song sung by a juvenile bird is referred to as the subsong. It is commonly likened to the babbling of human infants. Eventually the song of a juvenile bird develops to closely resemble the song of its tutor. At this point the juvenile has become sexually mature and its song has become stable.
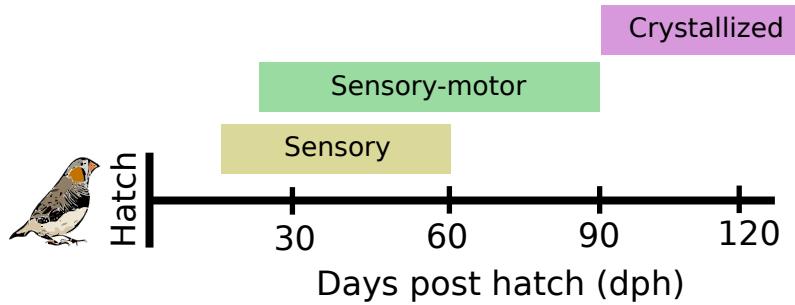
Figure 4.1: Timeline for song learning in zebra finch [8]. The sensory phase lasts from around 15 dph to 60 dph. The overlapping sensorimotor phase spans from around 25 dph to 90 dph. Approximately after 90 dph the song is crystallized.

There are three overlapping phases of song learning as illustrated in figure 4.1. The age of the juvenile birds is indicated in the units of days post-hatch (dph) which is essentially the number of days after hatching out of the egg. In the sensory phase (15 dph - 60 dph) the juvenile bird forms an internal representation of a tutor song to which its exposed. During the sensorimotor phase (25 dph - 90 dph), the juvenile starts to vocalize the song and tries to improve its song to be closer to the tutor song based on auditory feedback of its own singing. In this phase the juvenile's song is referred to as the plastic song. At around 90 *dph* the juvenile is left with a stable song which is similar to the tutor song. This is referred to as crystallization of the song.

4.2   Song system

In contrast to mammalian brain which is organized in cortical layers, songbird brain is organized in clustered neurons referred to as nuclei. A network of sensorimotor nuclei is collectively responsible for learning, producing and maintaining song. This network is referred to as the song system. A schematic representation of the connections between the nuclei of the song system is shown if figure 4.2.

HVC

Unlike the brain of mammals which is organized in cortical layers, the brain of songbirds is organized in clustered neurons referred to as nuclei. The analysis carried out in Chapter 5 focuses on structural changes in nuclei HVC, at the level of synapses. This nucleus is known to be involved in song learning and song production [56]. It is observed to be essential for the higher-level patterning of the song. Currently HVC is used as the proper name of this nucleus. Previously it was an acronym referring to a function which was later found out to be inconsistent with subsequent biological findings. HVC is the source of auditory input to the anterior forebrain pathway (AFP) and to the nuclei of descending motor pathway. It has two sets of efferent neurons projecting to RA and Area X. It has possibly four different sets of afferent neurons projecting from nucleus
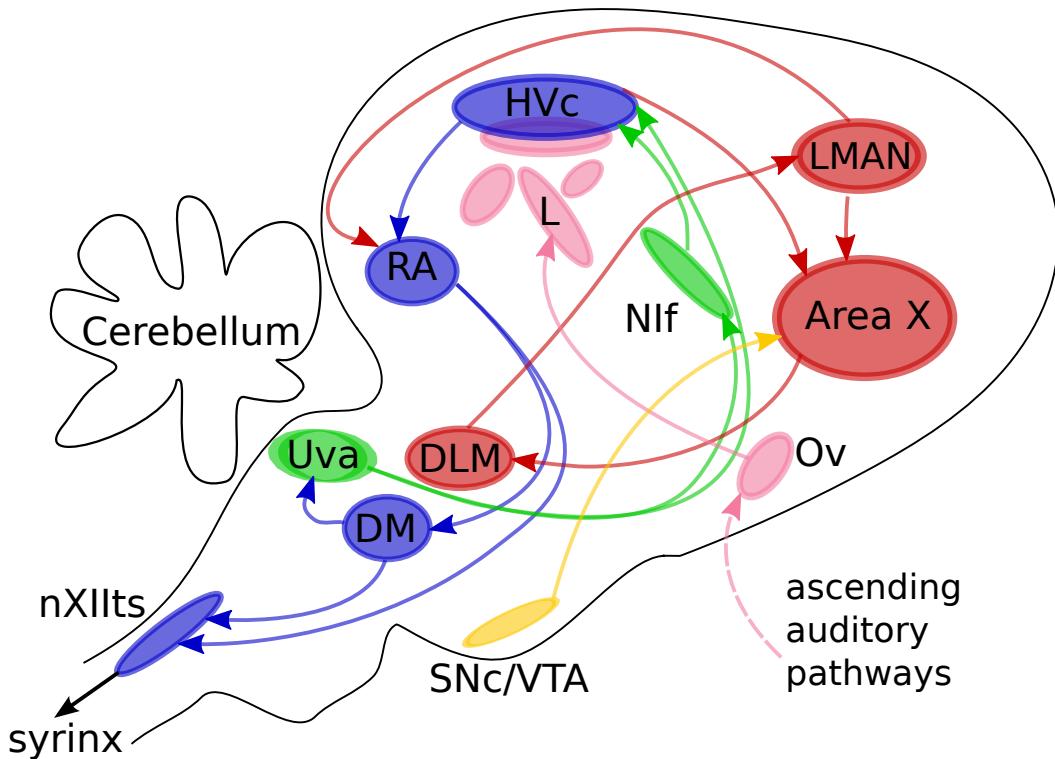
Figure 4.2: Generalized sagittal section of songbird brain highlighting neural pathways of vocal learning. The color code distinguishes between different pathways: blue - motor pathway, red - anterior forebrain pathway (AFP), pink - auditory pathway.

interfacialis (NIf), nucleus uvaeformis (Uva), median part of the magnocellular nucleus of the anterior neostriatum (mMAN) and field L (primary auditory telencephalon).

It has been discovered that neurons projecting from HVC to RA drives the singing of a crystallized adult song. The motor commands necessary for song production seem to be mainly present in the network containing HVC and RA. The activity of HVC to RA projection neurons encodes the temporal structure of the song in a highly sparse manner with significant stereotipy. In summary, the role of HVC neurons can be seen as providing sparse and stereotyped code to drive downstream motor commands [30].

Dendritic spine turnover rate in HVC neurons was observed to be affected by tutor exposure suggesting a direct link between song learning and synaptic changes in HVC neurons [66]. In the analysis carried out in the next chapter, we focus on the ultrastructural changes taking place in HVC at synapse level caused by different levels of exposure to tutoring.

Chapter 5

Effects of sensory exposure on MSB formation

5.1 Introduction

Since the early 1950s, electron microscopy (EM) has been a way of getting important insights of the anatomy and important structural properties of the nervous system. This was mainly because of the very high resolutions that could be obtained with EM compared to other microscopy techniques [70, 60]. At such high resolutions we are able to observe brain ultra-structures such as chemical synapses as shown in figure 5.1. This chapter elaborates an EM image driven analysis of some of the effects in zebra finch brain ultrastructure caused by sensorimotor learning.

5.1.1 Background

Synapses are the junctions through which information is passed from one neuron to another or to other types of cells. The message transmitted originates from the presynaptic partner and it is received by the postsynaptic partner. They are of two types: chemical synapses and electrical synapses. Electrical synapses are also called gap junctions. A gap junction is an electromechanical conductive link of ion channels between two neighbouring neurons. The more common form of synapses are chemical synapses which communicate using neurotransmitters packaged into vesicles that reside near the presynaptic membrane. In this chapter we only consider chemical synapses. Figure 5.1 illustrates how chemical synapses are seen in an EM image. On the presynaptic membrane a thick dark structure which is the presynaptic endocytic zone can also be observed. postsynaptic membranes have receptors on which the transmitted vesicles can dock and release neurotransmitters into the postsynaptic partner. These receptors are typically seen as the dark region on top of the membrane of the post-syanptic side and is referred to as the postsynaptic density. Synaptic vesicles of the presynaptic partner and the postsynaptic densities are clearly observable in EM images as shown in Figure 5.1.

Synapses can be broadly categorized into two types: asymmetric synapses (Gray's type 1) and symmetric synapses (Gray's type 2) [28]. In asymmetric synapses, the postsynaptic density is seen to be thicker than the presynaptic density as shown by EM images. This characteristic component of an asymmetric synapse is a fuzzy electron
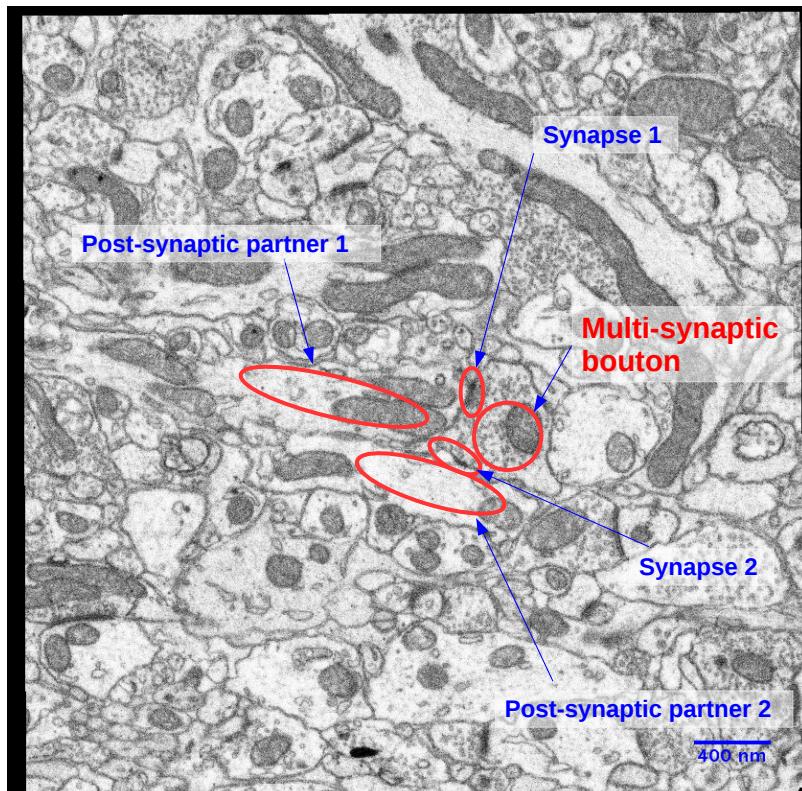
Figure 5.1: One section out of a FIBSEM image volume obtained from a sample of zebra finch HVC. This image is of size $5\mu m \times 5\mu m$. Each pixel corresponds to a $5nm \times 5nm$ area in the sample volume. In this image a multi-synaptic bouton is labelled along with its two synapses and the corresponding post-synaptic partners.

dense structure placed just beneath the membrane. This structure was found out to contain a variety of receptors, scaffolding proteins and signaling complexes involved in synaptic transmission and neuroplasticity [71]. Asymmetric synapses are commonly associated with excitatory junctions found mainly on dendrites and dendritic spines. EM image analysis further shows that the presynaptic boutons of asymmetric synapses contain spherical vesicles with a diameter of around 35 nm.

EM images of symmetric synapses which are considered to be inhibitory junctions show thinner postsynaptic densities than those of asymmetric synapses. Symmetric synapses are commonly found at the neuronal soma, along dendritic shafts and at the axon hillock. EM images further reveal that the vesicles in the presynaptic partner of a symmetric synapse contains vesicles with both spherical and flattened shapes.

When the synaptic junction is between two neurons, the postsynaptic partners are either dendritic spine heads or spine necks. A dendritic spine is the protrusion from a dendrite of one neuron towards a synaptic bouton of another neuron that receives input from that synaptic bouton. In a chemical synapse, dendritic spines express glutamate receptors on their surface for receiving neurotransmitters. A synaptic bouton with more than one postsynaptic partner is referred to as a multi-synaptic bouton (MSB). A postsynaptic partner can either be a dendritic shaft or spine.
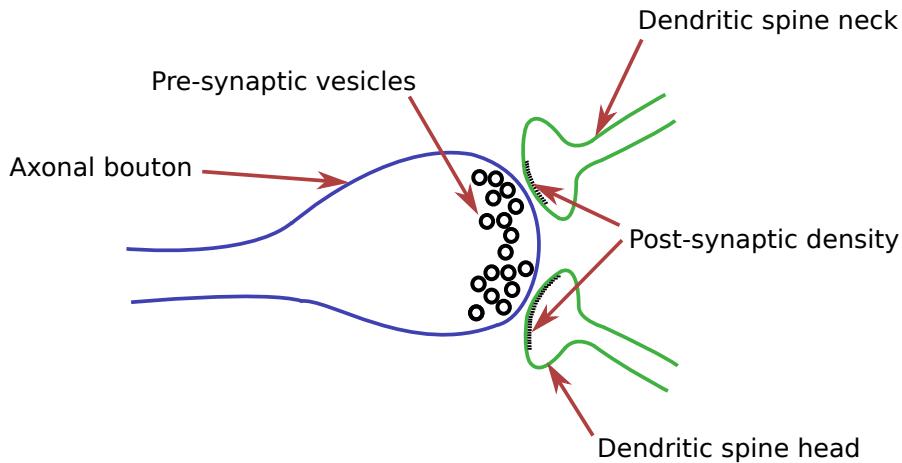
Figure 5.2: Simplified 2-dimensional illustration of a multi-synaptic bouton. Axonal bouton shown here has two postsynaptic partners.

5.1.2   Neuro-anatomical changes and learning

Learning of behaviors including motor skills depends on the brain's ability to respond to instructive experience. In 1890s, Ramon y Cajal and Eugenio Tanzi hypothesized for the first time that changes to the structure of the brain may occur as a consequence of experience [62, 79]. This important and fascinating capability of the brain which is believed to be based on changes in neural circuitry, is commonly referred to as 'plasticity'. Such changes usually involve the generation of new synaptic connections and modification of existing connections between dendrites and axons of different neurons. Recent studies have been able to confirm that dendritic spines are highly plastic structures that are observed to be a significant facilitator of learning and memory [34, 3, 69, 89].

Furthermore, synaptic strengthening which is thought to play an important role in learning has been observed to correlate with an increase in stability and the number and size of dendritic spines [4, 12, 14, 16, 52, 53, 88, 90]. In other experiments investigating effects of Long-term Potentiation (LTP) which is a form of synaptic plasticity, outgrowth of new spines or filopodia has been a consistent result of LTP-inducing stimulation [16, 52, 45].

5.1.3   Experience-dependent learning and MSBs

Proportion of MSBs compared to total number of axonal boutons has been quantified in different brain areas in rat and mice in previous studies. In the barrel cortex of adult mice over 1 month and it was found that the number of synapses on boutons varied from one to three. Only a small fraction of boutons were observed to be multi-synaptic (14%) [48]. It was also observed that the MSBs were generally larger than the single-synapse boutons (SSBs). Several studies have been carried out suggesting a relationship between experience dependent learning and the formation of multi-synaptic boutons. In cat visual cortex, 60% of the axonal boutons are found to be MSBs that have two postsynaptic partners that are usually dendritic spines [21]. In hippocampal area CA1 of rat, the following observations have been made regarding the proportions of different

types of connectivity of vesicle containing axonal boutons to dendritic spines [33]:

- 75% of the boutons connect to one dendritic spine (SSB)

- 21% of the boutons connect to more than one dendritic spine (MSB)

- 4% of the boutons didn't have any postsynaptic partner

It has also been observed that presynaptic axonal boutons vary in structure from one brain region to the other.

Different aspects of MSBs relating them to learning and memory has been experimentally investigated in several studies especially since the early 1990s. In 1991, Friedlander *et. al.* [21] observed that MSBs in cat visual cortex are influenced by visual experience during development. K. M. Harris (1995) [32] proposed a model for the spread of LTP between potentiated and presynaptic partners which are MSBs.

Jones *et. al.* [44] investigated the changes of the proportion of MSBs in the layer IV of the visual cortex in adult rats in a controlled experiment designed to expose the animals to different levels of stimulation. One group of rats were kept in an individually caged standard laboratory environment for 60 days. Another group of rats were kept together for 60 days in an enriched environment with toys that were changed daily. The number of MSBs normalized for the number of neurons (using unbiased counting) were found to be significantly higher in the layer IV of the rats kept in the enriched environment.

Toni *et. al.* (1999) [80] carried out an analysis by inducing LTP in rat hippocampal organotypic cultures. They performed an EM analysis with a protocol to identify calcium that had accumulated in the dendritic spines as a result of neural stimulation. It was concluded that LTP induction resulted in a significant increase in the proportion of MSBs. 3D reconstructions has shown that the presynaptic partners of MSBs frequently belong to the same dendrite. It was also observed in this study that pharmacological blockade of LTP prevented those morphological changes.

Knott *et. al.* in [48] investigated synapse formation and spine formation in an experience-dependent learning paradigm. This was carried out on transgenic mice that were subject to whisker trimming. Dendrites and spines were imaged live using two-photon laser scanning microscopy and ssEM was used to resolve synapses afterwards. It was observed that although spines tend to lack synapses soon after formation, the spines that persisted a few days always had synapses. It was further observed that new spines preferentially made synaptic connections with axonal boutons that already had other synapses. In addition to this work, [37] also provides evidence to suggest that new spines tend to form synaptic connections with large axonal boutons that already bear at least one synapse.

A novel form of rewiring plasticity was suggested in [50] where synapses with multiple contacts were formed within hours. It was observed that an axon of a pyramidal neuron approaches the dendrites of every neighboring pyramidal neuron multiple times without any structural bias towards the ones that end up forming synaptic connections. This observation suggested that the microcircuit could be in a state of all-to-all readiness so that it can rapidly change the synaptic connections without requiring the neurites to grow towards new targets.

Live imaging spanning over periods of 15 hours was used for the first time to investigate the dynamic behavior of MSBs by Reilly *et. al.* [65]. Neuronal tissue cultures

used in the experiment were obtained from rat hippocampus. For imaging, laser scanning confocal microscopy (LSCM) was used in combination with immunolabeling in order to simultaneously visualize pre- and postsynaptic partners. It was observed that the spines contacting MSBs are seemingly in competition with each other giving rise to some of the spines to become larger while the others disappear within tens of hours. This observation considered along with prior observations of MSB formation as a result of experience-dependent learning could suggest MSB formation as a transient phenomenon resulting from experience-dependent learning.

5.2   Objectives of EM image analysis

As described in the previous chapter, male juvenile zebra finches learn a stereotyped song from an adult male zebra finch. This process starts when the juvenile is around 15 days old and continues until it is 90 days old. During this period, sensorimotor learning of the juvenile gradually turns its babbling into a stereotyped song that closely resembles the tutor song.

The main objective of this chapter is to investigate effects on the formation of MSBs in juvenile zebra finch HVC at varying levels of exposure to a tutor. Brain area HVC of zebra finch song system is known to be responsible for vocal learning [30]. To study the anatomical changes we are interested in at synaptic resolution, we carried out a comparative analysis using nearly isotropic EM image volumes obtained from juvenile zebra finches that were subject to different tutoring durations.

With the analysis that follows we set out to answer the questions below:

1. Is there a significant difference of the number of MSBs between a group of juveniles that are tutored, and a group that is not tutored?

2. What changes can be observed in the number of MSBs when juvenile birds just start to getting tutored (after 1 day)?

3. What difference in the number of MSBs can be observed between a group of juveniles that just started getting tutored (for one day) and a group that has been tutored for a longer period of time (for 24 days)?

Along the lines of related work mentioned in the previous section which were mainly carried out on rodents, one hypothesis would be that the number of MSBs would be the highest for the juvenile zebra finches that just started to get tutored. This is inline with the observation in rodent experiments that:

- experience-dependent plasticity gives rise to an increase in the number of MSBs

- increase in MSBs is potentially a transient phenomenon where the number decreases over time

In summary the following hypothesis are to be tested:

1. The number of MSBs increase due to instructive experience

2. MSBs are seen as a transient state of newly formed synapses. Over time the the additional postsynaptic partners are eliminated [37, 57, 65].
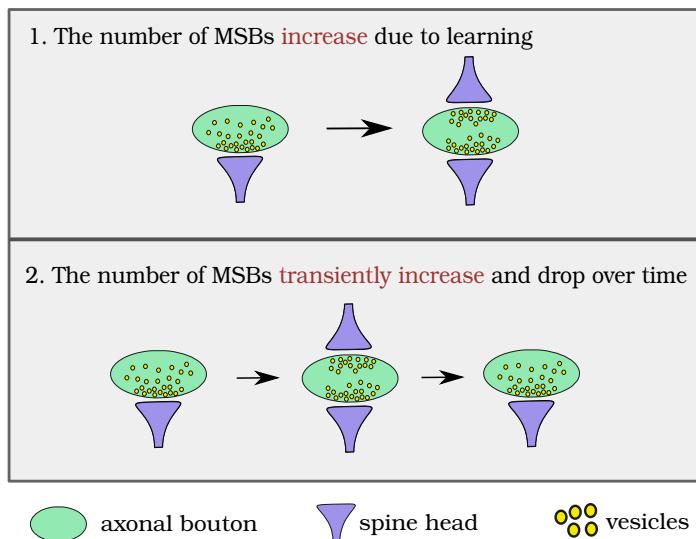
Figure 5.3: Summary of hypotheses on the effect of the number of MSBs due to learning effects that are to be tested for using an analysis of volumetric EM images.

5.2.1   Contributions

The experiments involving zebra finches were designed and carried out by Ziqiang Huang of the Institute of Neuroinformatics, ETH Zurich and University of Zurich. FIBSEM imaging was also performed by Ziqiang Huang. Leila Elabbady from Wellesley College, Wellesley, MA, USA assisted in the annotation, counting and proofreading of MSBs.

5.3   Methods

5.3.1   Experiment with three groups of juveniles

The experiment setup shown in Figure 5.4 was designed to have three different tutor exposure levels for three different groups of juvenile zebra finches of the same age. The three groups used in the experiment are:

- Isolate (ISO) - juveniles in this group did not receive any tutoring. Each of them are kept in isolation until they are 59 days old.

- Very late tutored (VL) - birds in this group receive tutoring just one day prior to the end of the experiment. On this day (day 59) each juvenile bird is housed with an adult male zebra finch to enable tutoring. Until such time each of the juveniles in this group is kept in isolation.

- Fully tutored (FT) - from day 35 until the end of the experiment (for 24 days), each bird in this group will be housed with an adult male zebra finch to be tutored.

Tissue samples of HVC of all the birds were obtained when they were 60 days old.
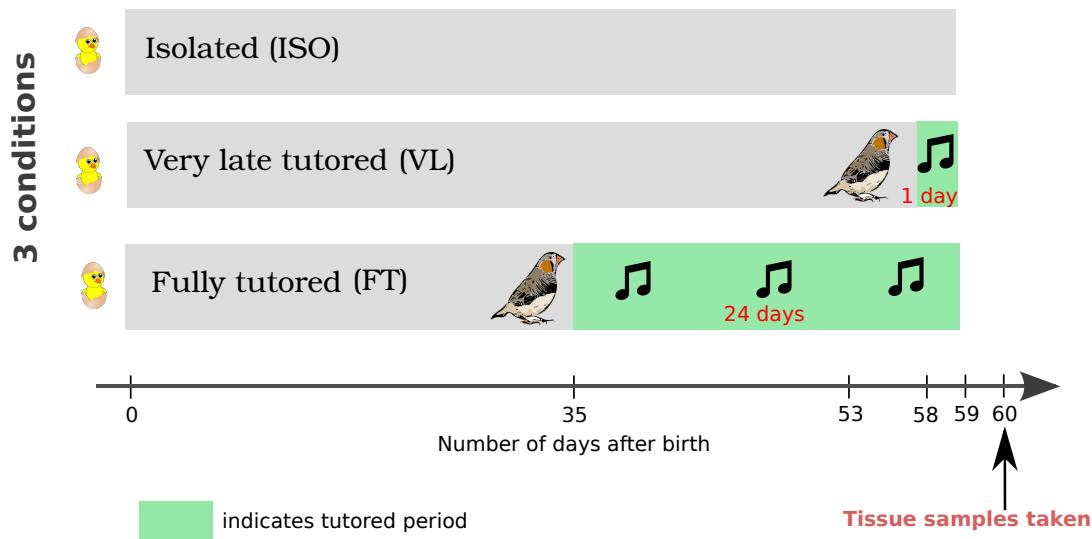
Figure 5.4: *Experiment setup:* Juvenile zebra finches of equal age are categorized in three different groups having four birds each. The experiment is carried out until each bird is 59 days old by which time the tissue samples from brain area HVC are taken. The birds in the first group (ISO) are kept in isolation and do not receive any tutoring during the course of the experiment. The second group (VL: very late tutored) will not be tutored until day 58 and on the 59th day each of those birds share a cage with a tutor. Each bird in the third group (FT: fully tutored) will share a cage with a tutor from day 35 for 24 days until perfusion is carried out at the end of the experiment.

5.3.2   Image acquisition

Tissue samples from HVC of each of the 12 birds in the three groups were obtained for imaging. In each case FIBSEM imaging was performed on a cube of dimensions $8\,\mu m \times 8\,\mu m \times 8\,\mu m$ resulting in nearly isotropic image volumes. The pixel size along the XY plane was $5\,nm \times 5\,nm$ and the thickness of each 'section' along the Z axis was set to $10\,nm$.

5.3.3   Identification of MSBs

Given the average thickness of a 'section' of a FIBSEM volume was around $10\,nm$, a typical MSB could span from 20 to 60 consecutive sections. The visual identification process of an MSB involves checking each bouton for having more than one postsynaptic partners. Therefore the process of manually annotating an MSB involves scrolling back and forth about each bouton and keeping track of the number of its postsynaptic partners.

The following steps describe how MSBs were manually annotated in FIBSEM volumes:

1. We start with overlaying synapse annotations on the volumetric image stacks. Although not compulsory, annotated synapses makes the task of browsing for synaptic boutons much faster.

2. Split each FIBSEM volume into two halves - one half to be annotated in the morning, the other in the afternoon by the same person on two different days. This is to lower any annotator bias.

3. Scroll through each sub-volume observing each pre-annotated synapse. For each synapse,

    (a) Identify the bouton corresponding to this synapse by inspecting the volume around it. Boutons are generally expected to be bulging out regions on an axon. However in certain cases the bulging is not significant. Another important feature of boutons is the presence of synaptic vesicle clouds.

    (b) Scroll back and forth along the identified bouton while counting the number of postsynaptic partners it has. Synaptic partners should be separate spine heads or dendritic shafts. We cannot be certain if two separate spine heads are part of the same dendrite because we do not have full 3D reconstructions of the dendrites around the MSBs.

    (c) Annotate the bouton as an MSB if it contains more than one postsynaptic partner

4. The following information is recorded for each MSB annotated:

    (a) Coordinates of MSB in the volume
    (b) Number of postsynaptic partners
    (c) Type of each synapse (symmetric or asymmetric)
    (d) Number of mitochondria in the MSB

5. Proofreading of the annotations is done by a different expert to lower annotator bias.

5.4   Results and discussion

MSBs are three dimensional structures that span several consecutive sections in an image volume. The visualizations of MSBs shown in Figures 5.5 and 5.6 were produced using a seeded region growing method [76] based on the graphcut [7] which is part of Ilastik [73].

5.4.1   Ambiguities in MSB identification

The main difficulty we faced during the identification of MSBs was in determining the 'boundary' of a bouton, especially when it is flatter, as opposed to being a clear protrusion of the axon. This in turn made the process of counting number of synapses per bouton difficult. As a result, deciding whether a set of synapses belonged to the same bouton or not was not immediately clear. However this was only the case for less than 10% of the MSBs annotated in each volume.

To minimize any effects caused by this ambiguity we focused on not only the protrusion of the bouton, but also the distance to synapses from its center. If a synapse is away from the center of bouton by a distance larger than twice the approximate radius at the center, that synapse would not be considered as part of this particular bouton.
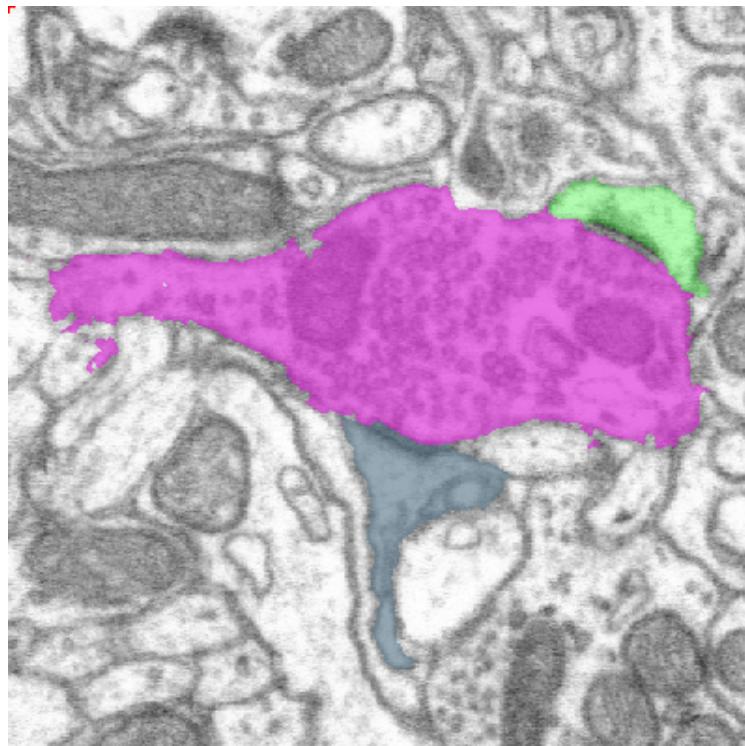
Figure 5.5: Annotation of a multi-synaptic bouton on a single EM section. The axonal bouton is annotated in magenta whereas the two dendritic spines are annotated in green and gray.

5.4.2   Number of MSBs

The main hypotheses that drove the analysis carried out in this chapter as described in section 5.2 was regarding the variation of the number of MSBs across different groups that receive different exposures to the tutors. Figure 5.8 shows the MSB counts normalized by volume averaged over each group. The normalized MSB count is observed to be higher in the very late tutored group (VL) than in both other groups. The normalized MSB count of the fully tutored group (FT) is in between that of the isolate group (ISO) and the very late tutored group. Kruskal Wallis One-way analysis of variance, there is not enough statistical evidence to conclude that the three groups have significantly different distributions of the normalized MSB counts (p = 0.4497). Therefore, it's not possible to claim that we observe a statistically significant transient increase of the MSB counts in the very late tutored group that drops when the birds are fully tutored, although the average normalized MSB counts tend towards such a pattern.

Figure 5.9 illustrates the normalized MSB counts for each tissue sample of all 12 birds in all three groups. A sub-volume analysis was performed for each volume by dividing each volume into four equal sub-volumes and counting MSBs in each sub-volumes to quantify the standard deviation of the normalized MSB count of each volume. The MSB counts for each of the sub-volumes are shown in Figure 5.10. As shown in Figure 5.9, the normalized MSB count does not seem to follow a significant trend within individual groups as captured by the Kruskal Wallis One-way analysis of variance.

(a)

(b)
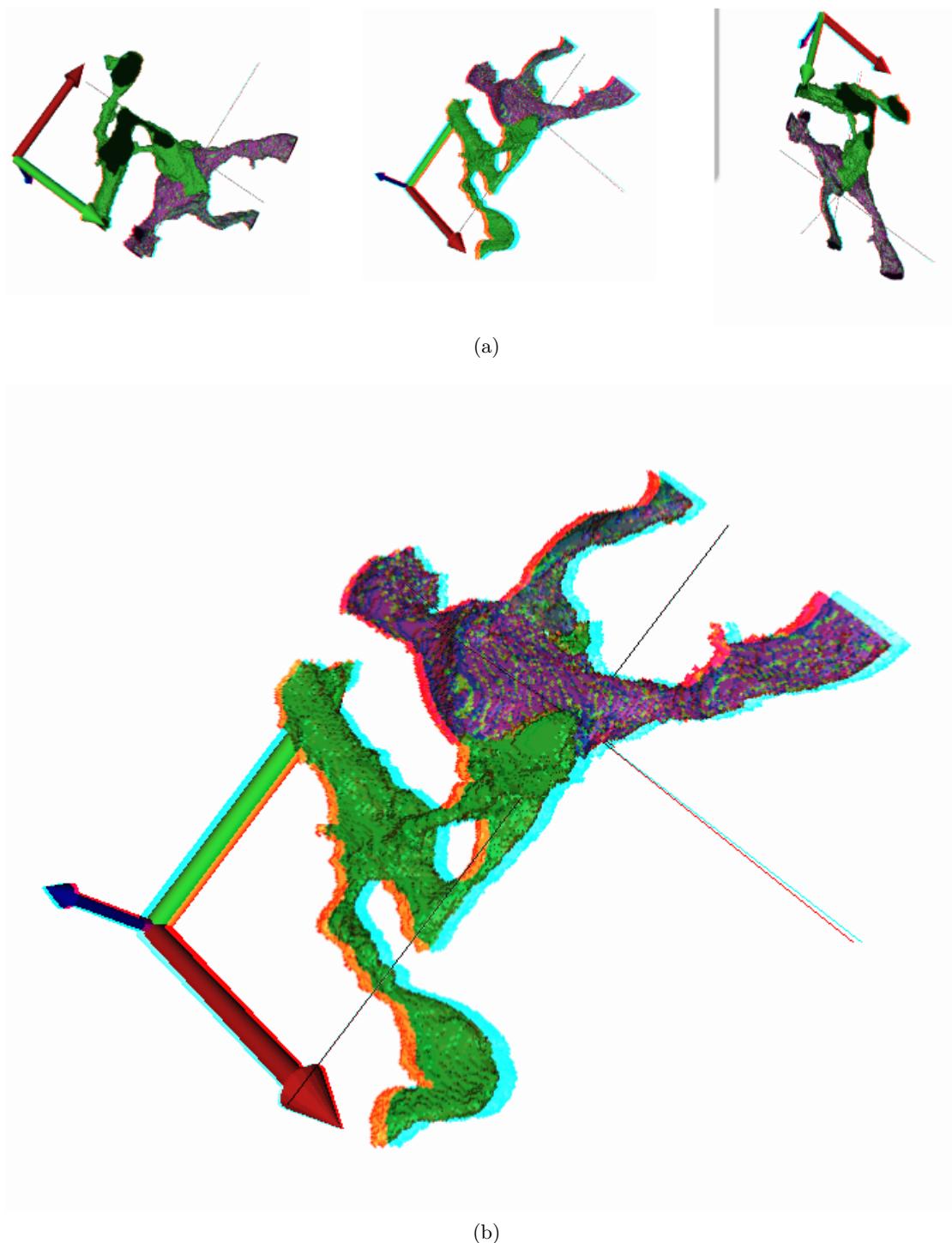
Figure 5.6: (a) Three different views of a 3D reconstruction of the same multi-synaptic bouton and postsynaptic partners shown in Figure 5.5. The axonal bouton is shown in magenta whereas the two dendritic spines are shown in green and gray. 3D visualization is possible with 3D glasses having separate red and blue channels. (b) Zoomed in 3D view of the same MSB.
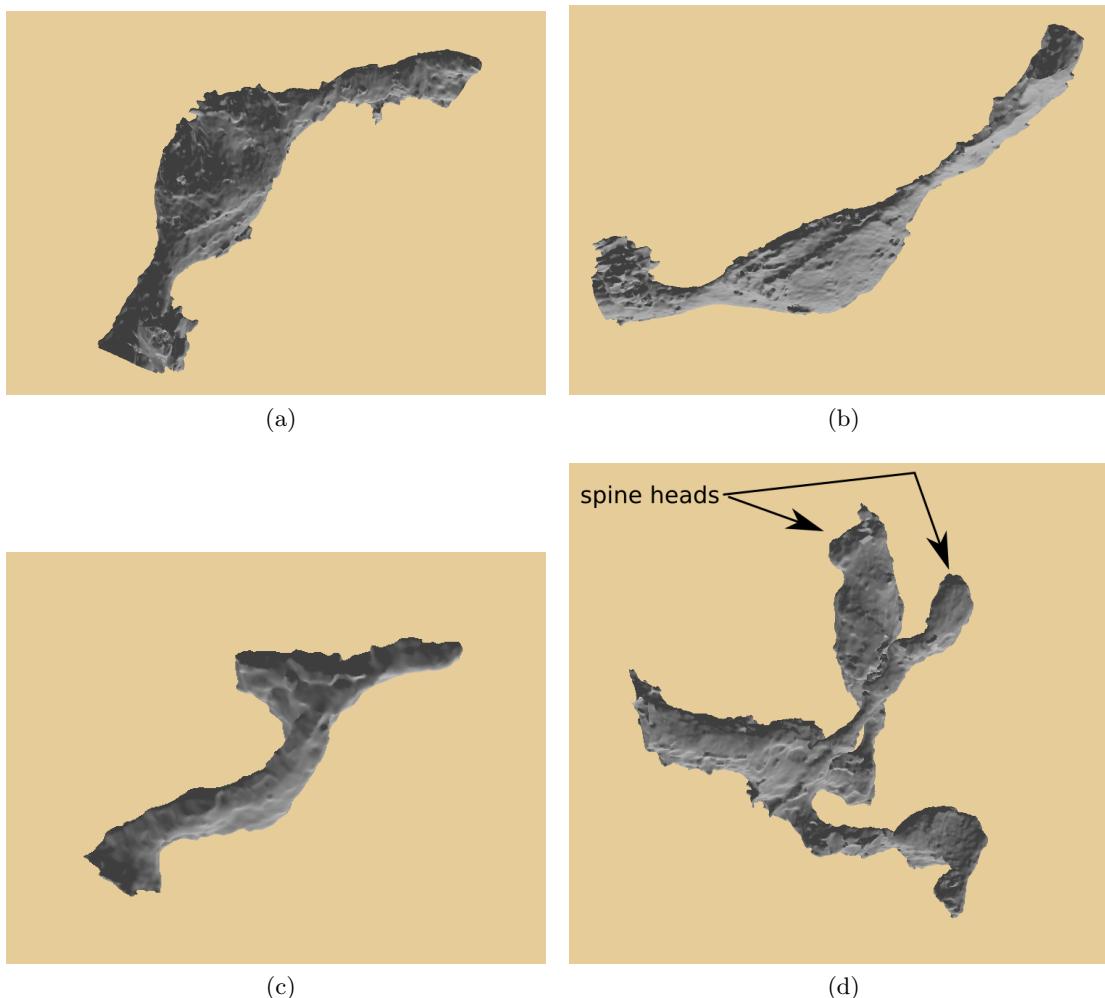
Figure 5.7: (a) and (b) show two different views of the 3D rendering of the axonal bouton shown in figure 5.6. 3D views of two different dendritic spines making synaptic connections with that bouton are shown in figures (c) and (d).

5.4.3   Number of synapses per MSB

Each MSB only contains synapses of one type: symmetric or asymmetric. We observed that the average number of synapses seen in symmetric MSBs were significantly higher than that of the asymmetric MSBs, with a p value of 0.000097346 from the Wilcoxon Rank-Sum test. This observation was not affected by the tutor exposure. The average number of synapses per symmetric MSB is close to 3.2 whereas it's around 2.2 for asymmetric MSBs as illustrated by the boxplot in Figure 5.12a. As shown in Figure 5.13, MSBs of both types mostly contain two synapses although occasionally there are MSBs with over five synapses.

(a)

(b)
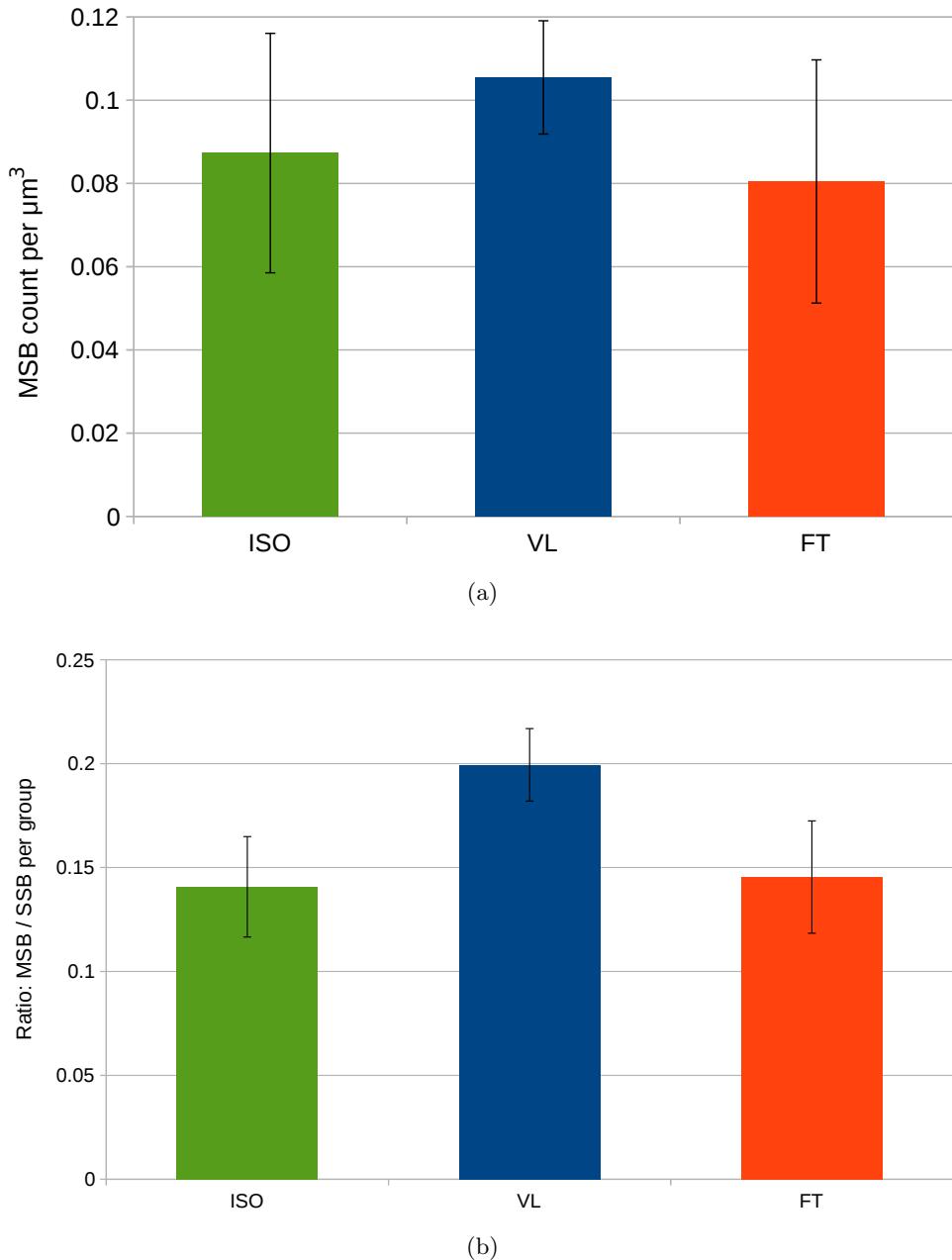
Figure 5.8: (a) Number of MSBs per µm³ averaged for each group. The one day tutored group (VL - very late tutored) has a slightly more mean normalized MSB count than other two groups. However, this difference was not found to be statistically significant. (b) Ratio of MSB to SSB counts averaged over each group, showing a similar pattern to normalized MSB counts. Standard deviation about the mean is denoted by the error bars.
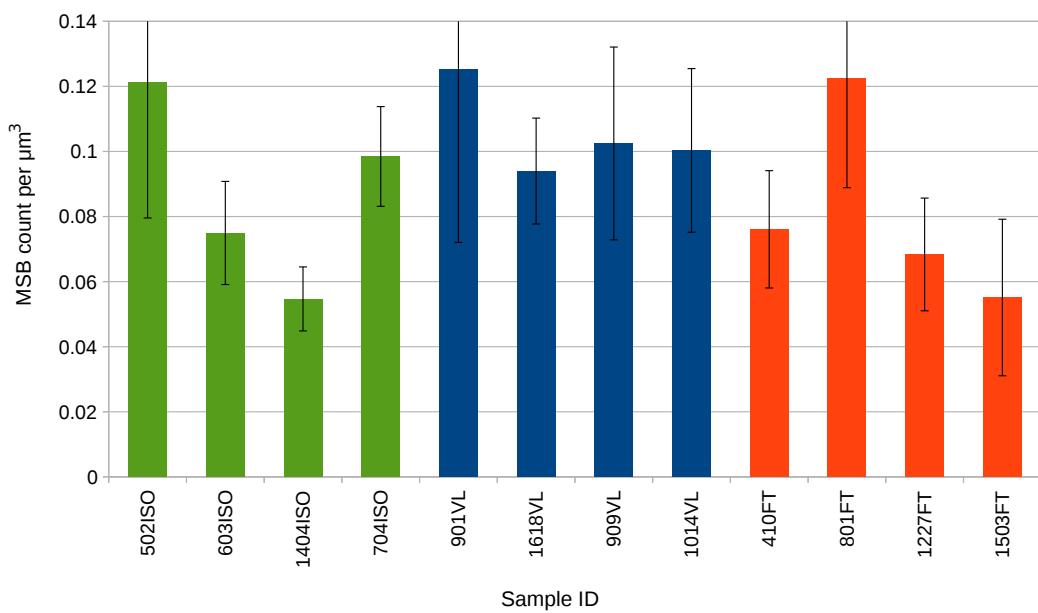
Figure 5.9: Bar graph showing number of MSBs per µm^3 for each volume, where each of them correspond to a different bird. The bars are color coded to show the group to which the bird belonged. Green: isolates (ISO), blue: very late tutored (VT), red: fully tutored (FT). The error bars correspond to the standard deviations obtained from the sub-volume analysis mentioned under 5.10
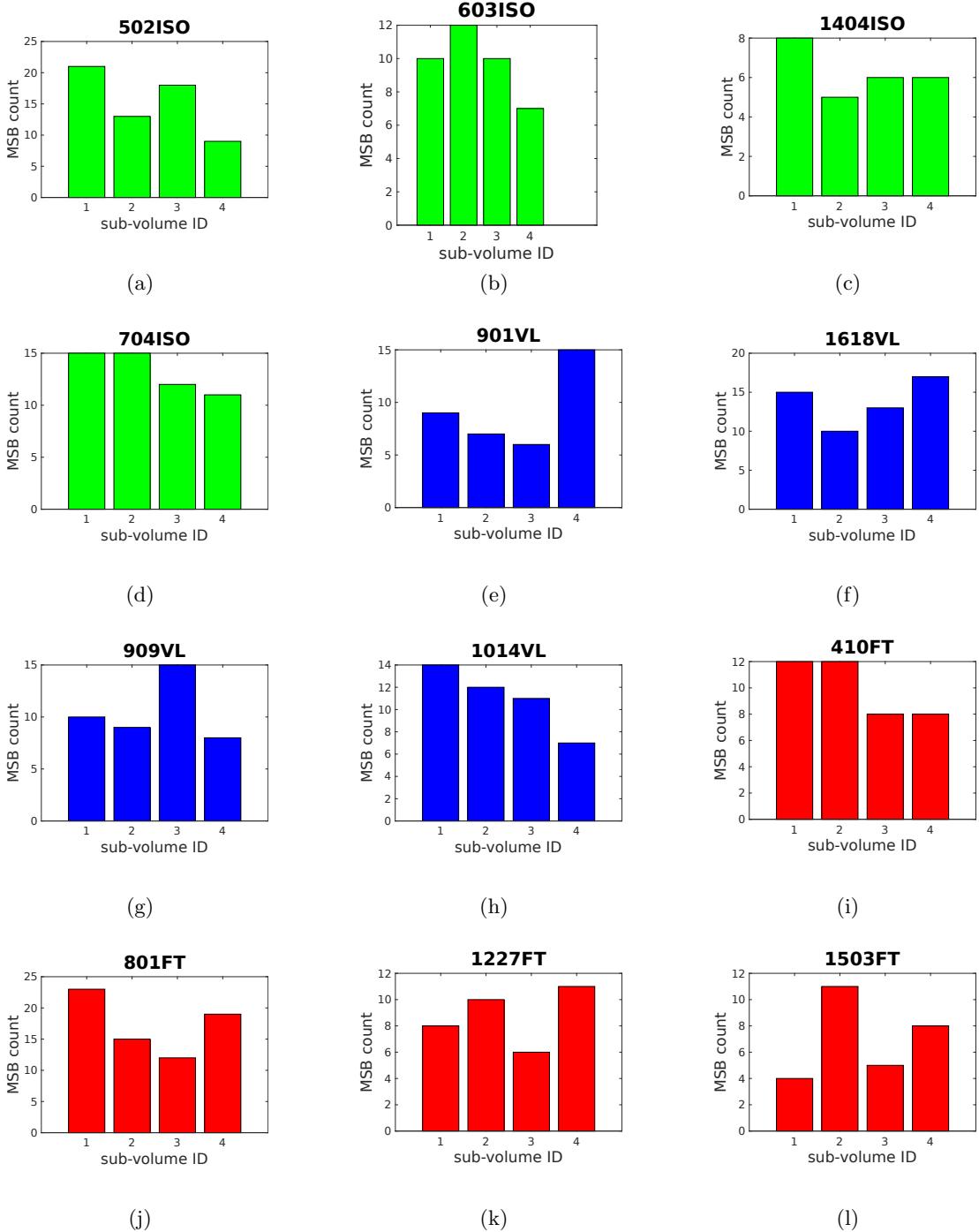
Figure 5.10: Each sub-plot shown in this figure corresponds to a sample that we analysed. The sub-plots corresponding to the isolate (ISO) group, the very late tutored (VL) group and the fully tutored (FT) group are shown in green, blue and red respectively. Each sub-plot has four bars that correspond to the number of MSBs found in each *sub-volume* of a sample. The sub-volumes are obtained by equally dividing the volume into 4. The MSB counts shown in these graphs are not normalized. Please refer to Figure 5.9 for a comparison of normalized MSB counts by volume.
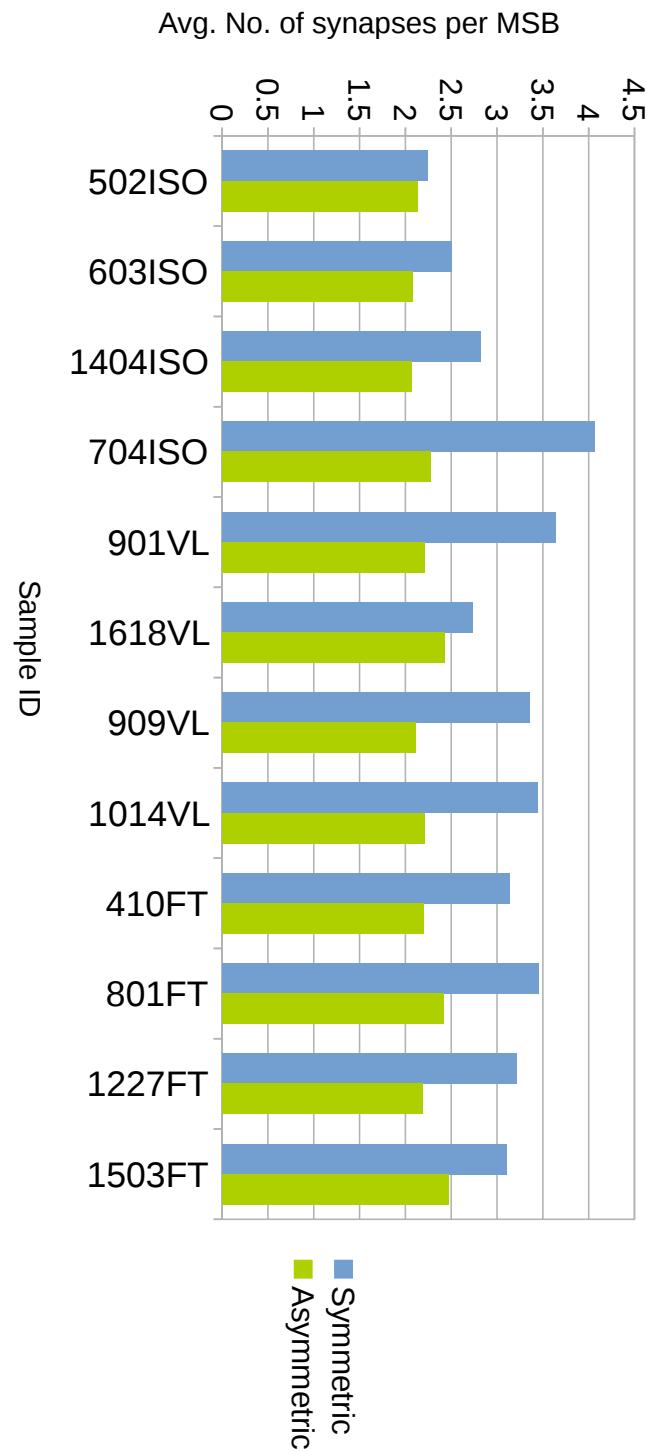
Figure 5.11: Average number of synapses per MSB in each volume. Blue bars represent MSBs with symmetric synapses whereas gree bars are for MSBs with asymmetric synapses. MSBs with symmetric synapses tend to have a higher number of synapses than MSBs with asymmetric synapses.

(a)

(b)
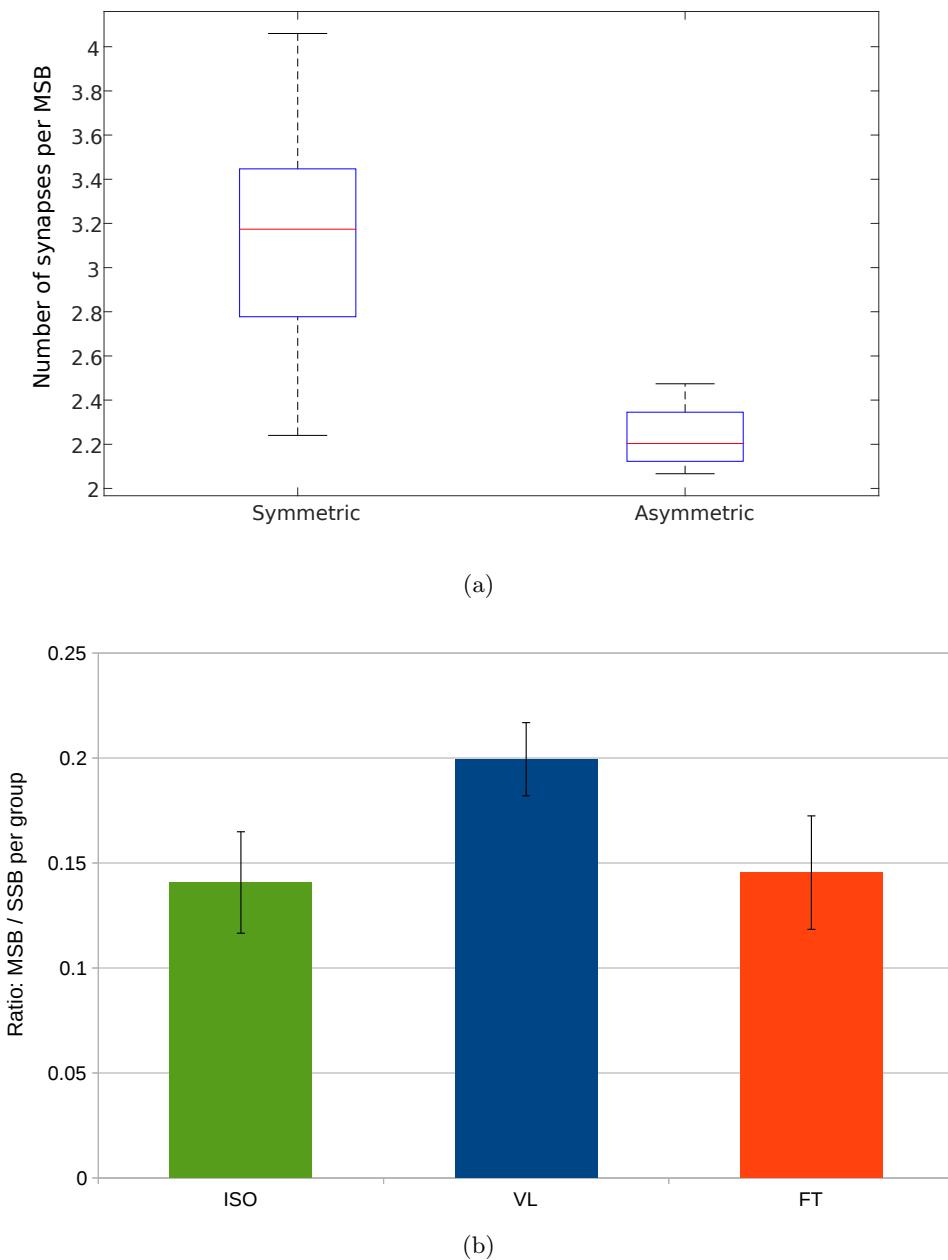
Figure 5.12: (a) Box plot showing average number of synapses per bouton over all the volumes. The average number of synapses in MSBs with symmetric synapses is significantly higher than MSBs with asymmetric synapses. (b) Ratio of MSB/SSB per each group.

(a)

(b)
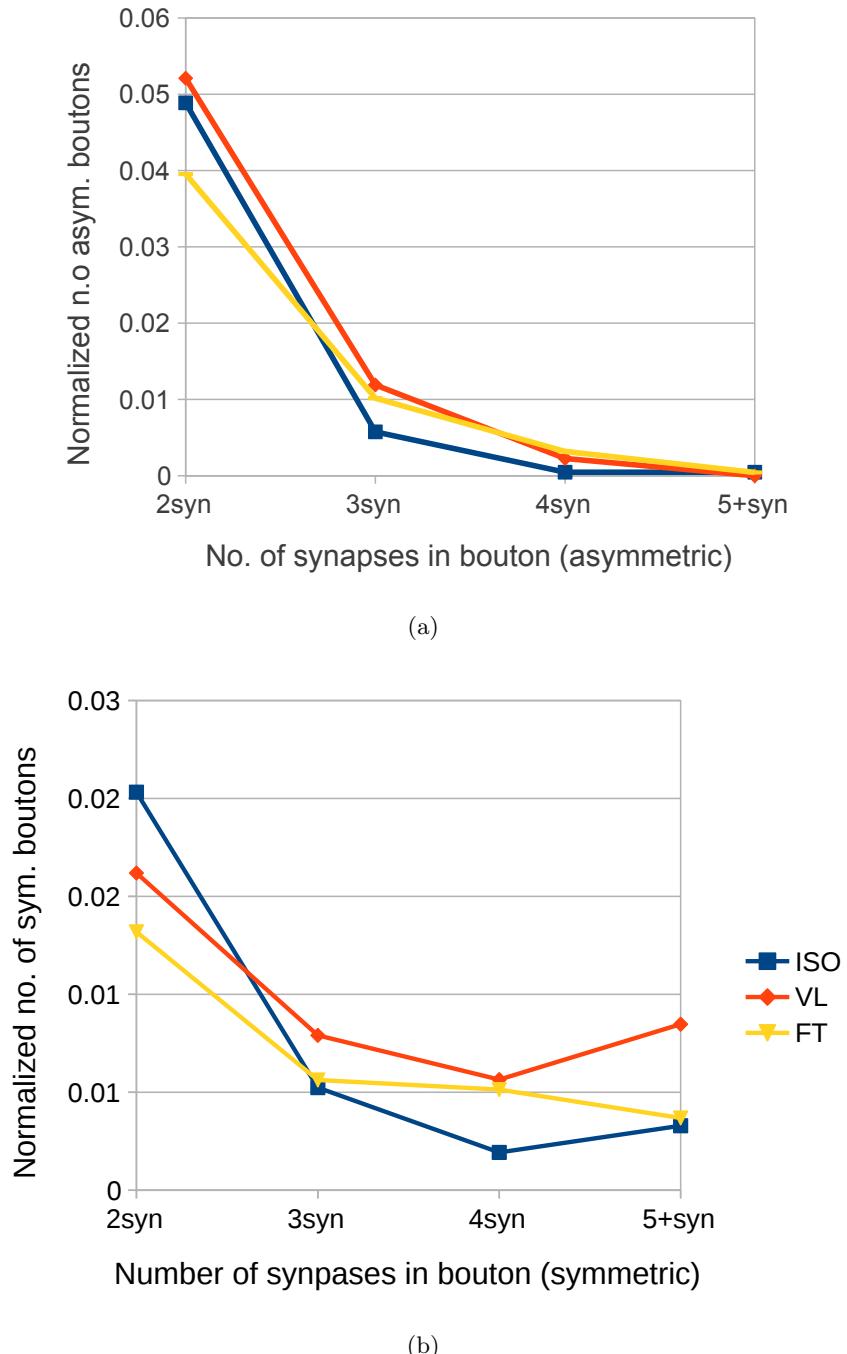
Figure 5.13: Number of boutons normalized by volume (per µm³) categorized by the number of synapses it has for: (a) asymmetric MSBs, (b) symmetric MSBs. Irrespective of tutor exposure and the type of MSB, most of them seem to have only two synapses although occasionally they could contain over five synapses especially in the case of symmetric MSBs.

5.5 Conclusion

Several previous studies performed on rat and mice provide evidence for an increase of MSBs due to experience-dependent learning. Later studies suggest that this increase is likely to be a transient phenomenon where the MSB numbers drop after an initial increase. Our analysis focuses on studying the variation of the number of MSBs in brain area HVC due to different levels of tutor exposure. HVC is known to play a crucial role in song learning and vocalization. We performed an image driven analysis where we obtained nearly isotropic 3D image volumes in the form of FIBSEM images from 12 juvenile birds that were kept in 3 different groups. On average, the group of birds that were tutored only one day prior to tissue samples being taken showed a slightly increased number of normalized MSB counts than the other two groups. One of the other groups was kept in complete isolation while the other group was exposed to a tutor for the last 24 days before the tissue samples were obtained. However, further analysis shows that the increase of the normalized MSB count was not statistically significant.

One important factor to be pointed out is that each of the volumes was of dimensions $8\,\mu m \times 8\,\mu m \times 8\,\mu m$. This approximately corresponds to one millionth of the average volume of zebra finch HVC. Therefore it's likely that much larger volumes of HVC tissue samples need to be analysed to make a stronger conclusion about the variation of the number of MSBs caused by tutor exposure.

Chapter 6

Discussion

6.1 Methods in EM-image-driven neuroscience research

Neuroscience research has greatly benefited from advances in imaging technologies and computer vision methods over the last few decades. Insights obtained from electron microscopy (EM) imaging has been instrumental in revealing neuroanatomical information at nanometer resolution where the presence of structures such as synapses can be observed.

EM-image-driven neuroscience research involves multiple steps starting from extracting biological tissue from an organism until the images are obtained, annotated and analysed for testing of biological hypotheses. The most important steps are as follows:

1. Extracting tissue

2. Sample preparation

3. Imaging using microscopes

4. Post-acquisition processing of images to adjust for distortions

5. Alignment of images

6. Manual/automatic/semi-automatic annotation of images

7. Performing statistical tests using annotated images

This dissertation spans stages 3, 4, 5, 6 and 7 and contributes with new methods for stages 4 and 6. For EM imaging, image alignment and statistical testing we used standard and state of the art methods available. This chapter provides a discussion of the methods used and developed along with results obtained during the entire dissertation.

6.2 Neuroimaging using EM

A variety of imaging methods are being used for obtaining images of biological tissue in neuroscience research. The focus of this dissertation only surrounds EM imaging that gives the highest possible resolutions achievable to date.

Traditionally, electron micrographs are visualized such that neuron membranes and synapses are shown in darker shades (closer to black) whereas cell interior is shown in

lighter shades (light gray). Having good levels of contrast consistently across an entire image facilitates more accurate image segmentation results using automated approaches. Conversely, erroneous segmentation results mostly correspond to areas where the SNR of a raw EM image is low. These problems are mainly caused by:

(a) Non-uniform heavy metal staining

(b) Thickness of sections being too high or too low

(c) Suboptimal parameters used in the microscope

(d) Imperfections of the microscope

For ssEM imaging, the thickness of the sections typically vary from 40 nm to 80 nm. In ssTEM imaging, the contrast is proportional to the amount of electrons trapped by the heavy-metal-stained structures of the specimen such as membranes and synapses. Therefore, if a section is too thin, the image contrast will be less because the specimen is not able to trap enough electrons. If a section is too thick, lateral displacement of neural processes along the Z axis of the section will result in a blurry EM image. On the other hand, ssSEM images are formed by backscattered or secondary electrons, instead of the electrons passing through the specimen. Therefore, as long as it is large enough the section thickness does not play a role in ssSEM image sharpness. However, if the thickness of a section is too low, it is likely that the specimen gets damaged or there is too little tissue to provide a sufficient signal.

While performing ssSEM, two of the main parameters for the microscope are the dwell time and the probe current. The dwell time is the duration of time taken by the electron beam to acquire a single pixel of the image. The probe current is the magnitude of the rate of flow of electrons (current) onto the specimen that eventually result in the backscattered and secondary electrons. The value of each of these two parameters should be high enough so that a good signal is obtained. However, if the values are too high the detected image will get saturated and the specimen will be destroyed. Furthermore, a higher dwell time results in a longer period of time taken for imaging. Therefore, by increasing the probe current, one could decrease the dwell time to retain the image quality while speeding up the imaging process. An optimal set of parameters is usually found by a trial and error process after starting off with a reasonable guess that matches the needs.

For the ssSEM images used for validating section thickness estimates in Chapter 2, a probe current of 400 nA and a dwell time of 0.7 μs was used. After setting these parameters, in order to acquire a good quality image the aperture, focus and astigmatism of the microscope have to be tuned. It's an iterative process where the tuning has to start at a low resolution and should be tuned repeatedly while increasing the resolution up to the desired level. This process can take a longer period of time for an inexperienced user of the EM equipment. That is mainly because the error that has to be optimized using set of three parameters is purely visual and the understanding of the direction of change of the error while varying each of the parameters needs some training. However, after a few attempts and some guidance it becomes a routine task. This is the main reason why the quality of the images acquired by a person has a steep improvement curve.

Performing SEM imaging on a specimen causes microscopic damages to the sample. Figure 6.1 is an AFM image of a specimen used for SEM imaging showing the depth of
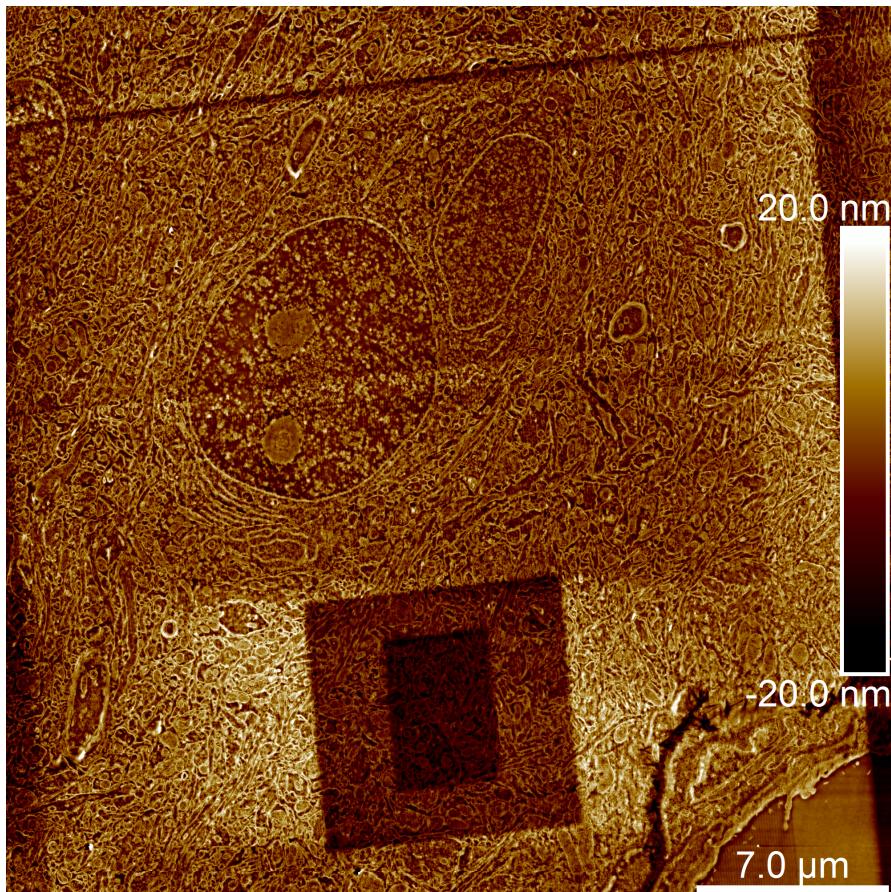
Figure 6.1: AFM image of a tissue section already imaged twice using ssSEM. The darker regions correspond to surfaces with lower height, which is due to the damage caused by the electron beam of the scanning electron microscope. The smaller region in the middle has been imaged twice with SEM. AFM measurements show that each time an SEM image is obtained, a layer of thickness around 20 nm of it is destroyed. *Image provided by Jozef Adamcik, D-HEST, ETH Zurich.*

the damage caused by the electron beam. We observed that each time an SEM image is obtained a layer of about 20 nm of the specimen is destroyed by the electron beam.

6.3   Image post-processing

Image alignment

Each section obtained by ssEM imaging is usually acquired as a set of overlapping image patches. Therefore, these image patches have to be stitched together and aligned with the rest of the sections to form a cohesive 3D image stack. The currently available solutions to the alignment problem which is also referred to as the registration problem does not have an exact solution. The best performing methods to date attempt to get an optimal alignment while minimizing the variation of image features across adjacent sections [68]. When an optimal solution is not achievable, the usual practice is to perform a pre-defined number of iterations of alignment adjustment that could hopefully result in a good enough solution. It is worth noting that there is no standard score to reflect the quality of an alignment. Therefore after applying an alignment routine, the user has to visually make a judgement of the alignment quality. When the results are not satisfactory the usual practice is to initialize the alignment with manual alignment points and apply the alignment method again. The manual alignment points are chosen by the user by following visual cues to describe the ono-to-one correspondence point in two adjacent images so that they are aligned as closely as possible.

While the EM image processing community has been significantly improving the alignment methods, one problem remains still unsolved. Especially with the advanced methods allowing to 'correct' for alignment of sub-parts of the same section by rotating and resizing, this could lead to a straightening of the neural processes giving an illusion of a high quality alignment. Although this could potentially benefit automatic 3D neural circuit reconstruction methods, it could lead to misrepresentation of the real data in terms of the absolute 3D arrangement. Regularization of the optimization task could be one solution to minimize such artefacts. However such effects are difficult to quantify and correct for, because there is no ground truth data set for comparison of alignment quality.

Anisotropy and thickness estimation

Sample preparation, cutting, alignment and imaging can cause anisotropy across the XY-plane and deviation of the section thickness from the desired value. Chapter 2 addresses this problem and provides an approach to estimate section thickness and anisotropy purely based on image statistics.

To our knowledge, there is no standard method for estimating image anisotropy for EM images. Furthermore, image segmentation methods usually assume isotropy along the XY-plane unless otherwise mentioned. The main use of the anisotropy estimate proposed in this work is to improve the thickness estimates, since the thickness estimation is done using the image statistics learned along the XY-plane for which we know the pixel resolution given by the microscope parameters. The primary use of a good thickness estimate is to have a more accurate quantification of the volume of tissue that is later used for any other volumetric analysis and to have more accurate spatial representations of 3D reconstructions derived from volumetric data.

Currently there is no standard dataset to compare the performance of section thickness estimation methods. As part of our work we have produced a dataset with directly

measured section thickness values using atomic force microscopy. Using this dataset
we have shown that our method produces better thickness estimates compared to other
methods that provide indirect thickness estimates for serial section EM images. Our
method is based on learning image statistics at pixel level using gaussian process regres-
sion which readily provides the variance of thickness estimates. Furthermore, we factor
out anisotropy of the XY plane from the image statistics so that the thickness estimates
made using the image statistics are more accurate.

However, the anisotropy calculation also includes the intrinsic anisotropy of the image
itself. This is usually a result of the rare occurrences where structures such as membranes
and mitochondria appearing to be aligned towards the same direction in an image. This
can cause imperfect thickness estimates for smaller images where the intrinsic anisotropy
could be dominant compared to average statistics. Therefore it's important to either
make sure the image size is large enough to not contain dominant intrinsic anisotropic
patterns in some of the images or remove such images by visual inspection so that they
won't affect the average statistics learned from the image stack.

6.4   Manual annotation and automatic segmentation of im-
ages

For connectomics research there are significant attempts to produce wiring diagrams for
entire brains of organisms or parts of it that possibly correspond to known brain func-
tions. Since these efforts require thousands of man hours when performed completely
manually, there is an immediate need for automation. Chapter 3 of the thesis addresses
part of the automation problem by providing a method to automatically segment neu-
ron slices in 2D EM sections. The main contribution of this chapter is the ability of
the proposed method to result in lower topological segmentation errors by incorporating
topological constraints into the segmentation task that results in pixel labels for neuron
membrane and individual 2D neuron slices. The topological constraints used here are all
local constraints that can be fully specified before the optimization is started. Interest-
ingly, these local constraints give rise to global behaviors such as closed loops defining
membranes around each neuron slice. If a method were to use global constraints, the
usual approach is to iteratively solve the segmentation problem, look for violation of
global constraints, add the violated constraints and resolve the problem.

Other than using this method for annotating 2D images, it can be potentially incorpo-
rated into 3D neuron reconstruction pipelines that depend on 2D segments. Furthermore,
it can also be used to produce over segmentations of 2D neuron slices that can be used
for 3D neuron reconstruction methods that are based on multiple 2D hypothesis [23].

Automatic segmentation methods that use membrane probability maps had a boost
in their performance during recent years due to the high quality probability maps that
can be generated using convolutional neural networks (CNN). However, CNNs demand
significant computational resources and require a large amount of training data which
could be expensive to produce especially because CNN models are not good at gener-
alising across different datasets unless the datasets have very similar pixel level image
statistics. Therefore, CNNs have to be trained for individual datasets requiring training
labels for each dataset. More recent work in this field has been attempting to perform

'transfer learning' where a network trained for one dataset can be used on a slightly different dataset by retraining only the last layer. This potentially requires much less training labels and also much less training time.

Automatic neuron segmentation methods directly benefit from any improvements made to the image acquisition process and pixel level classifiers such as CNNs. However, the accuracy of automated methods are still far away from being able to completely replace the human annotator. One possible suggestion to have an improvement over current segmentation approaches would be to focus on incorporating features from a broader scope so that split/merge errors can be lowered using a broader topological reasoning. A clear mathematical formulation for such an approach that can be efficiently solved using existing optimization schemes still remains open.

One other way to significantly improve the efficiency of the annotation process would be to have a combination of human and machine annotators to work in synergy. Such methods are traditionally referred to as semi-automatic methods. Machines are able to segment large volumes of data much faster than humans. If a method is able to speculate with high confidence where an automatic segmentation method might have made a mistake, a human annotator could correct it. In such a scenario, the human annotator would only look at tricky segmentation decisions where the machines might have failed, thereby significantly increasing the throughput ofimage annotation. Although this is not entirely a novel idea, such a method is yet to be successfully designed and deployed.

An important issue that cannot be separated from neuronal circuit reconstruction is the proofreading of annotations, which is generally performed manually. This results in a similar bottleneck in terms of efficiency as in manual image annotation. Therefore, improvements in proofreading methods would also greatly benefit neuronal circuit reconstruction efforts as well.

6.5   Hypotheses testing based on image data

Connectomics research thoroughly benefits from high throughput imaging and image annotation methods which are preferably automatic or semi-automatic. Therefore it requires a significant contribution from computer vision and machine learning method development. In fact, applications of imaging and image processing methods in neuroscience research go beyond investigations of neuron connectivity. Volumetric EM images can be used for quantitative and qualitative analysis of neuroanatomical structures such as synapses, synaptic vesicles, mitochondria, microtubules, axonal boutons, and dendritic spines. In Chapter 5 of this thesis an analysis of the variation of multisynaptic boutons (MSBs) as a result of sensorimotor learning is presented. The hypothesised transient change in the number of MSBs shortly after tutor exposure is not observed to be statistically significant, although a trend towards such a behavior could be seen. The analysis was performed using FIBSEM datasets, each of which corresponding to a $8 \times 8 \times 8 \, \mu m^3$ specimen of tissue taken from juvenile zebra finch HVC. Further analysis into this phenomenon should preferably use larger image volumes, for which either ssSEM or ssTEM has to be used instead of FIBSEM due to the specimen size limitations of the latter. However both ssSEM and ssTEM have less resolution in the third dimension of the image volume compared to FIBSEM which might render 3D reconstruction efforts more challenging.

Furthermore, an interesting observation could be made regarding the types of MSBs in all volumes analysed. MSBs having symmetric synapses (at inhibitory neurons) significantly outnumbered MSBs having asymmetric synapses (on excitatory neurons) irrespective of the experimental conditions.

6.6   Conclusion

The scope of this dissertation spans from EM image acquisition to development of computer vision methods to automatically analyze EM images, and applying image processing methods for biological hypothesis testing. Over the years, there have been many contributions from the EM image analysis community to different stages of the EM image analysis workflow. These methods mainly catered to post-acquisition processing and image annotation. Although significant progress has been made in recent years, there is still a plenty of room for improvement in terms of accuracy at every step of this pipeline. The methods developed as part of this dissertation contribute towards post-acquisition image processing and automated neuron segmentation that could potentially result in an overall improvement of the EM image analysis workflow.

Bibliography

[1] ISBI challenge 2012. `http://brainiac2.mit.edu/isbi_challenge/`. URL date: 2015-08-20.

[2] Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[3] Growth of dendritic spines: a continuing story. *Current Opinion in Neurobiology*, 15:67–72, 2005.

[4] V. Alvarez and B. Sabatini. Anatomical and physiological plasticity of dendritic spines. *Annual Reviews in Neuroscience*, 30, 2007.

[5] B. Andres, T. Kroeger, K. L. Briggman, W. Denk, N. Korogod, G. Knott, U. Kothe, and F. A. Hamprecht. Globally optimal closed-surface segmentation for connectomics. *ECCV*, 2012.

[6] G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Physical review letters*, 56(9), 1986.

[7] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*, 23(11):1222–1239, 2001.

[8] M. Brainard and A. Doupe. What songbirds teach us about learning. *Nature*, 417:351–358, 2002.

[9] A. Cardona, S. Saalfeld, S. Preibisch, B. Schmid, A. Chen, J. Pulokas, P. Tomancak, and V. Hartenstein. An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. *PLoS Biology*, 8(10), 2010.

[10] A. Cardona, S. Saalfeld, J. Schindelin, I. Arganda-Carreras, S. Preibisch, M. Longair, P. Tomancak, V. Hartenstein, and R. Douglas. Trakem2 software for neural circuit reconstruction. *PLoS one*, 7, 2012.

[11] F. Chen, P. Tillberg, and E. Boyden. Expansion microscopy. *Science*, 2015.

[12] D. Chklovskii, B. Mel, and K. Svoboda. Cortical rewiring and information storage. *Nature*, 431, 2004.

[13] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in Neural Information Processing Systems*, 2012.

[14] M. De Roo, P. Klauser, and D. Muller. LTP promotes a selective long term stabilization and clustering of dendritic spines. *PLoS Biology*, 2008.

[15] W. Denk, J. Strickler, and W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248:73–76, 1990.

[16] F. Engert and T. Bohhoeffer. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature*, 399, 1999.

[17] J. Fiala. Reconstruct: a free editor for serial section microscopy. *Journal of Microscopy*, 218:52–61, 2005.

[18] J. Fiala and K. Harris. Cylindrical diameters method for calibrating section thickness in serial electron microscopy. *Journal of Microscopy*, pages 468–72, 2001.

[19] E. Fischer, B. Hansen, V. Nair, F. Hoyt, and D. Dorward. Scanning electron microscopy. *Current Protocols in Microbiology*, 2012.

[20] J. Frank. *Electron Tomography*. Springer, 2nd edition, 2008.

[21] M. J. Friedlander, K. A. Martin, and D. Wassenhove-McCarthy. Effects on monocular visual deprivation on geniculo-cortical innervation of area 18 in cat. *Journal of Neuroscience*, pages 3268–3288, 1991.

[22] P. Fua and G. Knott. Modeling brain circuitry over a wide range of scales. *Frontiers in Neuroanatomy*, 2015.

[23] J. Funke, B. Andres, F. A. Hamprecht, A. Cardona, and M. Cook. Efficient automatic 3d-reconstruction of branching neurons from em data. *Proc. of the IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1004–1011, 2012.

[24] J. Funke, J. N. P. Martel, S. Gerhard, B. Andres, D. C. Cirean, A. Giusti, L. M. Gambardella, J. Schmidhuber, H. Pfister, A. Cardona, and M. Cook. Candidate sampling for neuron reconstruction from anisotropic electron microscopy volumes. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 17–24, 2014.

[25] J. Funke, F. Moreno-Noguer, A. Cardona, and M. Cook. Ted: A tolerant edit distance for segmentation evaluation. *arXiv:1503.02291*, 2016.

[26] J. Goldstein, D. Newbury, P. Echlin, D. Joy, C. Lyman, E. Lifshin, L. Sawyer, and J. Michael. *Scanning electron microscopy and X-ray microanalysis*. Springer Science+Business Media, 2003.

[27] F. Gonda, V. Kaynig, R. Thouis, D. Haehn, J. Lichtman, T. Parag, and H. Pfister. Icon: An interactive approach to train deep neural networks for segmentation of neuronal structures. *arXiv:1610.09032*, 2016.

[28] E. Gray. Axosomatic and axodendritic synapses in the cerebral cortex. *Journal of Anatomy*, 93:420–433, 1959.

[29] I. Gurobi Optimization. Gurobi optimizer version 6.5. houston, texas. *www.gurobi.com*.

[30] R. Hahnloser, A. Kozhevnikov, and M. Fee. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 2002.

[31] P. Hanslovsky, J. Bogovic, and S. Saalfeld. Post-acquisition image based compensation for thickness variation in microscopy section series. *International Symposium on Biomedical Imaging (ISBI)*, 2015.

[32] K. M. Harris. How multiple-synapse boutons could preserve input specificity during an interneuronal spread of LTP. *Trends in Neurosciences*, 18:365–369, 1995.

[33] K. M. Harris. Structural and functional organization of the synapse. *Springer Science & Business Media*, 2008.

[34] Y. Hayashi and A. K. Majewska. Dendritic spine geometry: functional implication and regulation. *Neuron*, 19:529–532, 2005.

[35] K. Hayworth, N. Kasthuri, R. Schalek, and J. Lichtman. Automating the collection of ultrathin serial sections for large volume tem reconstructions. *Microscopy and Microanalysis*, 12, 2006.

[36] M. Helmstaedter, K. Briggman, and W. Denk. High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nature Neuroscience*, 14:1081–1088, 2011.

[37] A. Holtmaat and K. Svoboda. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience*, 2009.

[38] H. Horstmann, C. Koerber, K. Saetzler, D. Aydin, and T. Kuner. Serial section scanning electron microscopy (s 3 em) on silicon wafers for ultra-structural volume imaging of cells and tissues. *PLoS One*, 7, 2012.

[39] V. Jain, H. S. Seung, , and S. C. Turaga. Machines that learn to segment images: a crucial technology for connectomics. *Curr. Opin Neurobiol.*, 20:653–66, 2010.

[40] V. Jain, S. Turaga, K. Briggman, M. Helmstaedter, W. Denk, and S. Seung. Learning to agglomerate superpixel hierarchies. *Adv. in Neural Information Processing Systems (NIPS)*, 2011.

[41] W. Jeong, J. Beyer, M. Hadwiger, R. Blue, C. Law, A. Vzquez-Reina, R. Reid, J. Lichtman, and H. Pfister. Ssecrett and neurotrace: interactive visualization and analysis tools for large-scale neuroscience data sets. *IEEE Computer Graphics and Applications*, 30:58–70, 2010.

[42] W. Jeong, J. Beyer, M. Hadwiger, A. Vzquez, H. Pfister, and R. Whitaker. Scalable and interactive segmentation and visualization of neural processes in em datasets. *IEEE Transactions on Visualization and Computer Graphics*, 15:1505–1514, 2009.

[43] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[44] T. Jones, A. Klintsova, V. L. Kilman, A. Sirevaag, and W. Greenough. Induction of multiple synapses by experience in the visual cortex of adult rats. *Neurobiology of Learning and Memory*, 68:13–20, 1997.

[45] P. Jourdain, K. Fukunaga, and D. Muller. Calcium/calmodulin-dependent protein kinase ii contributes to activity-dependent filopodia growth and spine formation. *Journal of Neuroscience*, 23:10645–10649, 2003.

[46] N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. A. Conchello, S. Knowles-Barley, D. Lee, A. Vzquez-Reina, V. Kaynig, T. R. Jones, M. Roberts, J. L. Morgan, J. C. Tapia, H. S. Seung, W. G. Roncal, J. T. Vogelstein, R. Burns, D. L. Sussman, C. E. Priebe, H. Pfister, and J. W. Lichtman. Saturated reconstruction of a volume of neocortex. *Cell*, 162:648–661, 2015.

[47] V. Kaynig, A. Vazquez-Reina, S. Knowles-Barley, M. Roberts, T. R. Jones, N. Kasthuri, E. Miller, J. Lichtman, and H. Pfister. Large-scale automatic reconstruction of neuronal processes from electron microscopy images. *Medical Image Analysis*, 11(1):77–88, 2015.

[48] G. Knott, A. Holtmaat, L. Wilbrecht, E. Welker, and K. Svoboda. Spine growth precedes synapse formation in the adult neocortex in vivo. *Nature Neuroscience*, 9:1117–1124, 2006.

[49] G. Knott, H. Marchman, D. Walland, and B. Lich. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *The Journal of Neuroscience*, page 2959 2964, 2008.

[50] J.-V. Le B and H. Markram. Spontaneous and evoked synaptic rewiring in the neonatal neocortex. *Proceedings of the National Academy of Sciences of the USA*, 103:13214–13219, August 2006.

[51] H. Li, R. Zhao, and X. Wang. Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification. *ArXiv e-prints arXiv:1412.4526*, 2014.

[52] M. Maletic-Savatic, R. Malinow, and K. Svoboda. Rapid dendritic morphogenesis in ca1 hippocampal dendrites induced by synaptic activity. *Science*, 283, 1999.

[53] M. Matsuzaki, N. Honkura, G. Ellis-Davies, and H. Kasai. Structural basis of long term potentiation in single dendritic spines. *Nature*, 429, 2004.

[54] M. Meila. Comparing clusterings - an information based distance. *Journal of Multivariate Analysis*, 98:873–895, 2007.

[55] F. Meyer. Topographic distance and watershed lines. *Elsevier Signal Processing*, pages 113–125, 1994.

[56] R. Mooney. Neurobiology of song learning. *Current Opinion in Neurobiology*, 19:654–660, 2009.

[57] U. V. Nagerl, G. Kostinger, J. C. Anderson, K. A. Martin, and T. Bonhoeffer. Protracted synaptogenesis after activity-dependent spinogenesis in hippocampal neurons. *Journal of Neuroscience*, 2007.

[58] J. Nunez-Iglesias, R. Kennedy, J. S. T. Parag, and D. Chklovskii. Machine learning of hierarchical clustering to segment 2d and 3d images. *PLoS ONE*, 8(8), 2013.

[59] S. Ogawa, T. Lee, A. Kay, and D. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of National Academy of Science, USA*, pages 9868–9872, 1990.

[60] S. Palay and G. Palade. The fine structure of neurons. *J Biophys. Biochem. Cytol. 1*, pages 69–88, 1955.

[61] L. Peachey. Thin sections. a study of section thickness and physical distortion produced during microtomy. *Journal of Biophysical and Biochemical Cytology*, pages 233–42, 1958.

[62] S. Ramn y Cajal. Consideraciones generales sobre la morfologa de la clula nerviosa. *La Veterinaria Espaola*, 37, 1894.

[63] W. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 1971.

[64] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning.* The MIT Press, Cambridge, MA,, 2006.

[65] J. E. Reilly, H. H. Hanson, M. Fernandez-Monreal, S. L. Wearne, P. R. Hof, and G. R. Phillips. Characterization of msb synapses in dissociated hippocampal culture with simultaneous pre- and postsynaptic live microscopy. *PLoS ONE*, October 2011.

[66] T. Roberts, K. Tschida, M. Klein, and R. Mooney. Rapid spine stabilization and synaptic enhancement at the onset of behavioural learning. *Nature*, 463, 2010.

[67] S. Saalfeld, A. Cardona, V. Hartenstein, and P. Tomancak. Catmaid: collaborative annotation toolkit for massive amounts of image data. *Bioinformatics*, 25:1984–1986, 2009.

[68] S. Saalfeld, R. Fetter, A. Cardona, and P. Tomancak. Elastic volume reconstruction from series of ultra-thin microscopy sections. *Nature Methods*, 9:717–720, 2012.

[69] M. Segal. Dendritic spines and long-term plasticity. *Nature Reviews Neuroscience*, 6:27–284, 2005.

[70] S. Seung. *Connectome: How the Brain's Wiring Makes Us who We are.* Houghton Mifflin Harcourt, 2012.

[71] M. Sheng and E. Kim. The postsynaptic organization of synapses. *Cold Spring Harb Perspect Biol.*, 12, 2011.

[72] J. Small. Measurement of section thickness. *Fourth European Congress on Electron Microscopy*, 1968.

[73] C. Sommer, C. Straehle, U. Koethe, and F. A. Hamprecht. ilastik: Interactive learning and segmentation toolkit. *Eighth IEEE Int. Sym. on Biomedical Imaging (ISBI)*, pages 230–233, 2011.

[74] J. Sporring, M. Khanmohammadi, S. Darkner, N. Nava, J. Nyengaard, and E. Jensen. Estimating the thickness of ultra thin sections for electron microscopy by image statistics. *International Symposium on Biomedical Imaging (ISBI)*, 2014.

[75] D. Sterio. The unbiased estimation of number and size of arbitrary particles using the disector. *Journal of Microscopy*, pages 127–136, 1984.

[76] C. Straehle, U. Koethe, G. Knott, and F. Hamprecht. Carving: Scalable interactive segmentation of neural volume electron microscopy images. *Proceedings of MICCAI*, 6891:653–660, 2011.

[77] A. Suissa-Peleg, D. Haehn, S. Knowles-Barley, V. Kaynig, T. Jones, A. Wilson, and H. Pfister. Automatic neural reconstruction from petavoxel of electron microscopy data. *Microscopy and Microanalysis*, 22:536–537, 2016.

[78] Y. Takahashi, T. Chiba, H. Sameda, S. Ohtori, M. Kurokawa, and H. Moriya. Organization of cutaneous ventrodorsal and rostrocaudal axial lines in the rat hindlimb and trunk in the dorsal horn of the spinal cord. *Journal of Comparative Neurology*, 445, 2002.

[79] E. Tanzi. I fatti e le induzione nell'odierna istologia del sistema nervoso. *Riv. Sper. Freniat. Med. Leg.*, 19:419–472, 1893.

[80] N. Toni, P.-A. Buchs, I. Nikonenko, C. Bron, and D. Muller. LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature*, pages 421–425, 1999.

[81] F. Tschopp, J. N. P. Martel, M. C. S. C. Turaga, and J. Funke. Efficient convolutional neural networks for pixelwise classification on heterogeneous hardware systems. *13th IEEE Int. Sym. on Biomedical Imaging (ISBI)*, pages 1225–1228, 2016.

[82] A. Vazquez-Reina, M. Gelbart, D. Huang, J. Lichtman, E. Miller, and H. Pfister. Segmentation fusion for connectomics. *IEEE International Conference on Computer Vision (ICCV)*, pages 177–184, 2011.

[83] A. Vazquez-Reina, D. Huang, M. Gelbart, J. Lichtman, E. Miller, and H. Pfister. Segmentation fusion for connectomics. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.

[84] C. K. I. Williams. *Prediction with Gaussian processes: From linear regression to linear prediction and beyond*, pages 599–621. The MIT Press, Cambridge, MA,, 1999.

[85] C. K. I. Williams and C. E. Rasmussen. Gaussian processes for regression. *Advances in Neural Information Processing Systems 8*, pages 514–520, 1996.

[86] M. Xu, T. Jarrell, Y. Wang, S. Cook, D. Hall, and S. Emmons. Computer assisted assembly of connectomes from electron micrographs: Application to caenorhabditis elegans. *PLoS one*, 8, 2013.

[87] P. Yushkevich, J. Piven, H. Hazlett, R. Smith, S. Ho, J. Gee, and G. Gerig. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31:1116–1128, 2006.

[88] R. Yuste and T. Bonhoeffer. Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annual Reviews in Neuroscience*, 24:1071–1089, 2001.

[89] R. Yuste and T. Bonhoeffer. Genesis of dendritic spines: insights from ultrastructural and imaging studies. *Nature Reviews Neuroscience*, 5:24–34, 2004.

[90] K. Zito, V. Scheuss, G. Knott, T. Hill, and K. Svoboda. Rapid functional maturation of nascent dendritic spines. *Neuron*, 61, 2009.

[91] K. Zuiderveld. *Contrast limited adaptive histogram equalization*. Graphics gems IV, Academic Press Professional, Inc., 1994.