

DEEP LEARNING METHODS FOR 3D  
SEGMENTATION OF NEURAL TISSUE IN EM  
IMAGES

BENJAMIN EISNER

A THESIS

PRESENTED TO THE FACULTY  
OF PRINCETON UNIVERSITY  
IN CANDIDACY FOR THE DEGREE  
OF BACHELOR OF SCIENCE IN ENGINEERING

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

COMPUTER SCIENCE

ADVISER: SEBASTIAN SEUNG

MAY 5, 2017

© Copyright by Benjamin Eisner, 2017.

All rights reserved.

# Abstract

Current state-of-the-art methods in the task of performing 3D segmentation of EM stacks typically rely on a multi-stage processing of input data. Roughly, data processing consists of: acquisition, realignment, preprocessing, representation transformation, and post-processing. While most techniques strive to be fully automatic in each of these stages, errors inevitably occur. When they occur, they must either be manually corrected, or the errors will inevitably propagate through the rest of the pipeline to the detriment of the output segmentation. In this paper, we will explore various methods of improving accuracy at two of the stages: representation processing, and realignment. Specifically, we explore deep-learning based approaches that maximize representation processing performance on well-aligned data, and then explore various methods of learned realignment to make processing robust to misalignment. We find that while hand-crafted alignment methods currently outperform learned alignment methods, the training results suggest that further exploration of more sophisticated learned realignment schemes could potentially outperform hand-crafted methods. Additionally, we release a modular segmentation framework, DeepSeg, that allows for automatic segmentation of EM datasets and provides a flexible way to experiment with different techniques.

# Acknowledgements

To my parents.

# Contents

Abstract . . . . .	iv
Acknowledgements . . . . .	v
List of Tables . . . . .	ix
List of Figures . . . . .	x
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Sub-Problems . . . . .	2
<b>2 Problem Background and Related Work</b>	<b>9</b>
<b>3 Models</b>	<b>11</b>
3.1 2D Boundary Detection . . . . .	11
3.2 3D Segmentation . . . . .	11
3.3 Misalignment . . . . .	12
<b>4 Implementation</b>	<b>13</b>
4.1 Software Design . . . . .	13

4.2	Hardware Design . . . . .	13
4.3	Deployment Design . . . . .	14
<b>5</b>	<b>Training</b>	<b>15</b>
5.1	Regularization . . . . .	15
5.2	Data Augmentation . . . . .	15
5.3	Model Training Behavior . . . . .	16
5.3.1	2D Boundary Detection . . . . .	16
5.3.2	3D Segmentation . . . . .	16
5.3.3	Misalignment . . . . .	16
<b>6</b>	<b>Evaluation Metrics</b>	<b>17</b>
6.1	Pixel Error . . . . .	17
6.2	Rand Score . . . . .	17
<b>7</b>	<b>Results</b>	<b>18</b>
7.1	2D Boundary Detection . . . . .	18
7.2	3D Segmentation . . . . .	18
7.3	Misalignment . . . . .	19
<b>8</b>	<b>Conclusion</b>	<b>20</b>
	<b>Bibliography</b>	<b>21</b>



# List of Tables

# List of Figures

- 1.1 An example of 2D boundary detection. Left: the original image taken with an electron microscope. This particular example is neuron tissue taken from *Drosophila melanogaster* in a dataset created for the ISBI 2012 EM segmentation challenge [3]. The resolution of each pixel is 4nm x 4nm. Right: The ground truth boundaries corresponding to cell membranes in the input image, as labeled by human experts. The labels are binary values, although the actual border deliniation is somewhat arbitrary due to the fact that real applications of boundary detection are invariant to small differences in boundary shapes. . . . . 3

1.2	An example of a 2D cross-section of a 3D segmentation. Left: one of the original images in a stack of images taken with an electron microscope. This particular example is neuron tissue taken from the common mouse in a dataset used in the ISBI 2013 EM segmentation challenge [5]. The resolution of each pixel is 6nm x 6nm, and each image represents a slice 30nm thick. Right: The ground truth segmentation corresponding to a segmentation of each individual object in the input image, as labeled by human experts. The labels are unique identifiers, although the border deliniation is somewhat arbitrary due to the fact that real applications of boundary detection are invariant to small differences in boundary shapes. . . . .	5
1.3	An example of a 3D stack of EM images that contains a misalignment. Left: The provided alignment of a stack. This represents a misalignment where the fourth image in a stack of images actually represents a slice slightly translated in position. Right: The correct alignment of the stack, where all the pixels in the fourth position have been translated enough such that the structures depicted in the input data line up in the z-direction. . . . .	8

# Chapter 1

## Introduction

### 1.1 Overview

In the field of computational neuroscience, there exist a class of problems relating to mapping the neuron-level structure of the brain. For instance, one might want to precisely model the neural connectivity of an abnormal mouse brain, or observe the connectivity and topology of a worm at different stages of its development. Conventionally, the structure of the brain is inferred from images, whether they are thin slices of a brain imaged with an electron microscope, volumetric images acquired using digital radiography systems (i.e. fMRI, CAT, etc), or visible-spectrum video of exposed brain tissue. Although these imaging technique generate information at different resolution levels, they invariably present a huge data problem: when researchers are presented with small-scale image data, it is fundamentally infeasible to efficiently infer the connectivity and structure of a small cluster of neurons by hand,

let alone an entire brain or nervous system, simply because the amount of neurons in a brain is too large.

Many attempts have been made to automate the process of inferring connectivity and topology from images using various algorithmic and machine learning models. In the past five years or so, many of the most successful attempts at this class of problems have utilized Convolutional Neural Networks (CNNs) to achieve their high performance. The goal of this year-long project is to explore many of the different CNN-based approaches that have gained recognition in the past few years in several sub-problems, evaluate their performance and enumerate their deficiencies, and attempt to design new architectures that achieve improved performance in these sub-problems. In addition to increasing performance on established benchmarks, we also make contributions on new sub-problems for which there are no established benchmarks.

## **1.2 Sub-Problems**

The motivation for the research in this field is to better understand the connectivity of neural tissue. Since this is such a broad goal, it stands to reason that there a number of intermediary sub-problems that can be tackled to learn about connectivity. Several of the sub-problems have been heavily studied, and various public competitions have been organized that provide labeled training data and unlabeled test data, encouraging competitors to achieve maximum performance against a certain benchmark. As

we developed our models, we submitted their predictions to several of these open competitions, often performing well.

## 2D Boundary Detection for EM Images

One of the more established problems in the field is 2D Boundary Detection. The problem statement is such: given a 2-dimensional single-channel (i.e. greyscale) image of neural tissue taken with an electron microscope, produce an image that labels the boundaries of all the distinct cells in the image. An example of this boundary-detection task can be found in Figure 1.1. This task is made particularly difficult by the existence of organelles with well-defined borders, as well as blood vessels and structured interstitial tissue.

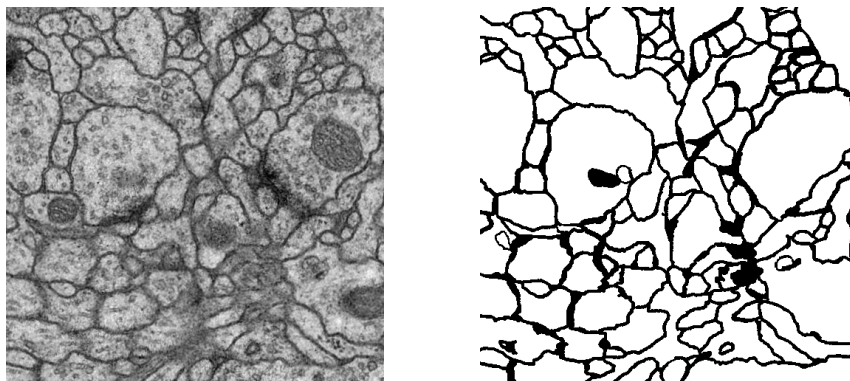


Figure 1.1: An example of 2D boundary detection. Left: the original image taken with an electron microscope. This particular example is neuron tissue taken from *Drosophila melanogaster* in a dataset created for the ISBI 2012 EM segmentation challenge [3]. The resolution of each pixel is 4nm x 4nm. Right: The ground truth boundaries corresponding to cell membranes in the input image, as labeled by human experts. The labels are binary values, although the actual border deliniation is somewhat arbitrary due to the fact that real applications of boundary detection are invariant to small differences in boundary shapes.

One prominent competition that evaluates performance on this sort of task is the International Symposium on Biomedical Imaging (ISBI) EM Segmentation Challenge, which has had active submission since 2012. The ISBI Challenge organizers provides a training set of EM images, along with a set of binary boundary maps. The challenge website describes the training data as “a set of 30 sections from a serial section Transmission Electron Microscopy (ssTEM) data set of the *Drosophila* first instar larva ventral nerve cord (VNC). The microcube measures 2 x 2 x 1.5 microns approx., with a resolution of 4x4x50 nm/pixel” [3]. This resolution description implies that each pixel represents a 4x4nm patch on the surface of a slice, with each slice being 50nm thick. We build prediction systems using several different architectures, regularization methods, and data transformation techniques. We make several submissions to the leaderboard, ultimately scoring quite competitively.

### 3D Segmentation for EM Volumes

More recently, the sub-problem of automatic 3D Segmentation has emerged as a popular research area. This sub-problem is formulated as such: given a stack of 2-dimensional EM images generated that represent a 3-dimensional volume of tissue (i.e. the images were taken of successive physical slices of tissue), produce a segmentation<sup>1</sup> of the set of images that uniquely labels each discrete entity in the original volume. That is, if a tissue volume contains a neuron that passes vertically through several different slices, then the portions of each slice through which the neuron

---

<sup>1</sup>A segmentation of an image or a stack of images is defined as producing a label for each pixel in the image or stack of images, where each unique label corresponds to a discrete object in the physical volume.

passes would be labeled with the same identifier. This problem is significantly more complicated than the boundary prediction problem stated before, because it requires an awareness of context in 3 dimensions, rather than 2. Additionally, most EM datasets are anisotropic, meaning that the resolution is not uniform in all directions (specifically, the z-direction perpendicular to the plane of each image is generally dilated). An example of a segmentation can be found in Figure 1.2.

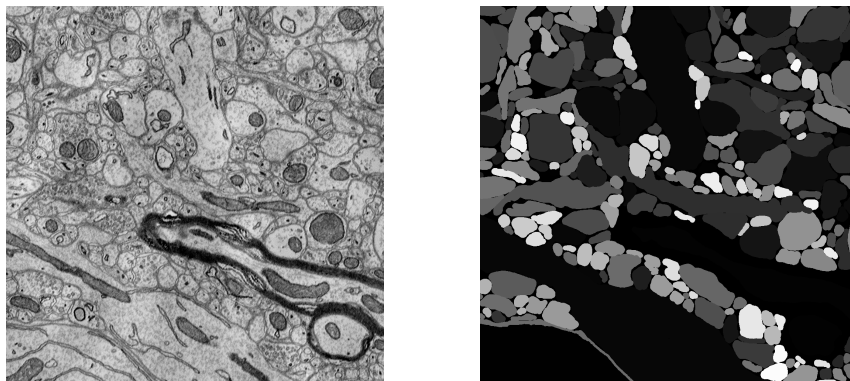


Figure 1.2: An example of a 2D cross-section of a 3D segmentation. Left: one of the original images in a stack of images taken with an electron microscope. This particular example is neuron tissue taken from the common mouse in a dataset used in the ISBI 2013 EM segmentation challenge [5]. The resolution of each pixel is 6nm x 6nm, and each image represents a slice 30nm thick. Right: The ground truth segmentation corresponding to a segmentation of each individual object in the input image, as labeled by human experts. The labels are unique identifiers, although the border deliniation is somewhat arbitrary due to the fact that real applications of boundary detection are invariant to small differences in boundary shapes.

Trivially, the complexity of objects in 3 dimensions is potentially much greater than in two dimensions, so it makes sense that any learning method used to train a system that performs segmentation might be adept at certain types of volumetric data, and inept at others. To evaluate methods on different types of volumetric data,



we selected two different challenges that provide us with samples of neural tissue that have different geometric properties, not only due to geometric differences in the underlying tissue but also because of differences in sample preparation techniques. These two challenges are the SNEMI3D Segmentation Challenge and the CREMI Segmentation Challenge.

The SNEMI3D Segmentation Challenge is a highly active challenge (organized in advance of ISBI 2013), and provides a stack of EM images for training, along with ground truth segmentations of the EM images in 3 dimensions. The challenge website describes the training and testing data as “stacks of 100 sections from a serial section Scanning Electron Microscopy (ssSEM) data set of mouse cortex. The microcube measures 6 x 6 x 3 microns approx., with a resolution of 6x6x30 nm/pixel” [2]. Like the ISBI 2012 dataset, the SNEMI3D dataset is anisotropic, and particularly dilated in the z-direction. Additionally, the data is from mouse cortex, rather than from *Drosophila*, and the geometry of the tissue is significantly different. **TODO:** Discuss our submissions to this leaderboard.

The Circuit Reconstruction from Electron Microscopy Images (CREMI) Challenge is a somewhat less-active challenge organized in advance of MICCAI 2016[4]. The challenge provides three datasets for training, all of which are volumetric samples of *Drosophila melanogaster*. The training and testing data are stacks of 125 sections from an ssSEM data set, with each slice having a resolution of 4x4x40nm/pixel. These datasets are also anisotropic, being dilated in the z-direction. Furthermore, the types of neurons sampled are quite diverse between datasets: from visual inspection, some of the neurites in one of the datasets is much thinner than those in

the others, suggesting that models might perform differently when trained/tested on these different datasets. Finally, these datasets are quite a bit noisier than ISBI or SNEMI3D: there are many more major misalignments, many patches of blur, and some slices are missing entirely. These datasets will provide a good measure of how robust our methods are to noise in volumetric data. **TODO:** Discuss our submissions to this leaderboard.

## Image Realignment

One major hurdle in inferring neural structure from EM images is that the image acquisition process is inherently noisy. While the EM imaging technologies used for the creation of neuron images (typically TEM) are quite stable, there is often variance in sample preparation techniques, resulting in all sorts of distortions and errors at imaging time. One particular type of error, image misalignment, occurs during the slicing of sample tissue, when some physical factor causes a resulting slice to be warped or translated in such a way that the resulting stack of images is misaligned. Intuitively, this means that every point in one EM slice data does not necessarily map to the point directly below it the neighboring slice. An example of slice misalignment can be visualized in Figure 1.3.

The problem of misalignment within a set of EM images particularly induces problems in the task of 3D Segmentation. While most techniques are rather invariant to small misalignments (particularly CNNs, which can be trained to be invariant to warping of many kinds), large misalignments can often induce false splitting in segmentations. Very deep CNNs trained with a really diverse set of data would

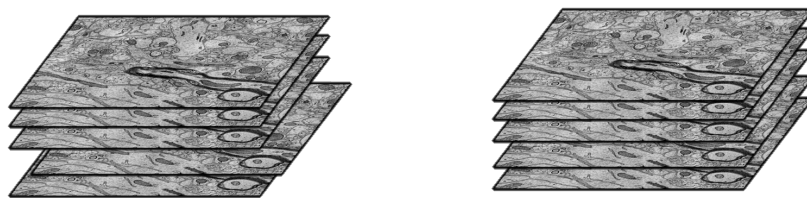


Figure 1.3: An example of a 3D stack of EM images that contains a misalignment. Left: The provided alignment of a stack. This represents a misalignment where the fourth image in a stack of images actually represents a slice slightly translated in position. Right: The correct alignment of the stack, where all the pixels in the fourth position have been translated enough such that the structures depicted in the input data line up in the z-direction.

likely be able to compensate for these sorts of misalignments, but it would be more prudent to develop a more efficient strategy for automatically healing misalignments in the data.

**TODO:** There are no current benchmarks or competitions for this task, but we will qualitatively (and quantitatively through downstream tasks) attempt to evaluate our methods on this sub-problem.

## Chapter 2

# Problem Background and Related Work

**TODO:** Flesh out more background on CNNs

The problem of determining the connectivity of a brain falls in the sub-field of connectomics, which has been a lively area for research for over 30 years. The first full connectome of an organism was created in 1986, producing the mapping of the brain of *C. elegans* [7]. Since then, partial and full topological and connectivity maps have been created on various organism, often using electron microscopy and careful hand-reconstruction to do so. More recently, by genetically modifying organisms to produce proteins that become phosphorescent in the presence of calcium (calcium is released across the synapse between a dendrite and an axon when a neuron is fires), researchers have been able to monitor both brain activity and neural structure using video photography in the visible spectrum [6].

In terms of the computational approaches to automating connectome reconstruction, many groups have attempted to create systems to accomplish near-human accuracy when labelling neurons in the brain. In the early 2010's, a group of researchers published an open dataset and created a global challenge to use machine learning methods to label neurons in image slices of a brain, resulting in the creation of models with near-human accuracy [3]. More recently, researchers at Princeton University applied similar approaches to automatically detecting neurons in videos formed with the calcium imaging techniques mentioned above [1].

The work in the field continues to improve mapping techniques, but our understanding of learning methods in both computational neuroscience and computer vision are in their infancy, and are not yet practical for deployment on a larger organism (i.e. a human) due to computational and accuracy issues. This project hopes to improve on existing methods, even if incrementally, so that we can better understand what an efficient mapping approach capable of mapping a full brain might look like.

# Chapter 3

## Models

### 3.1 2D Boundary Detection

- N4 and variants
- VD2D and variants
- U-Net and variants

### 3.2 3D Segmentation

- N4 and variants
- VD2D and variants
- U-Net and variants
- Flood Filling Networks

### 3.3 Misalignment

- Spatial Transformer Networks

# Chapter 4

## Implementation

### 4.1 Software Design

- Modular net definitions
- Data pipelining
- Learner interface
- Tensorflow

### 4.2 Hardware Design

- GPU Training with immense speedup



## 4.3 Deployment Design

- Docker w/ Nvidia-docker

# Chapter 5

## Training

### 5.1 Regularization

- Batch normalization
- Standardization
- Dropout

### 5.2 Data Augmentation

- Flipping/Rotating
- Mirroring
- Elastic Deformation

## **5.3 Model Training Behavior**

### **5.3.1 2D Boundary Detection**

- N4 and variants
- VD2D and variants
- U-Net and variants

### **5.3.2 3D Segmentation**

- N4 and variants
- VD2D and variants
- U-Net and variants
- Flood Filling Networks

### **5.3.3 Misalignment**

- Spatial Transformer Networks

# Chapter 6

## Evaluation Metrics

### 6.1 Pixel Error

### 6.2 Rand Score

# Chapter 7

## Results

### 7.1 2D Boundary Detection

- N4 and variants
- VD2D and variants
- U-Net and variants

### 7.2 3D Segmentation

- N4 and variants
- VD2D and variants
- U-Net and variants
- Flood Filling Networks

## 7.3 Misalignment

- Spatial Transformer Networks

## Chapter 8

## Conclusion

# Bibliography

- [1] Noah J. Apthorpe, Alexander J. Riordan, Rob E. Aguilar, Jan Homann, Yi Gu, David W. Tank, and H. Sebastian Seung. Automatic Neuron Detection in Calcium Imaging Data Using Convolutional Networks. jun 2016.
- [2] Ignacio Arganda-Carreras, H. Sebastian Seung, Ashwin Vishwanathan, and Daniel R. Berger Vishwanathan. ISBI 2013 challenge: 3D segmentation of neurites in EM images — SNEMI3D: 3D Segmentation of neurites in EM images, 2013.
- [3] Ignacio Arganda-Carreras, Srinivas C. Turaga, Daniel R. Berger, Dan Cirean, Alessandro Giusti, Luca M. Gambardella, Jürgen Schmidhuber, Dmitry Laptev, Sarvesh Dwivedi, Joachim M. Buhmann, Ting Liu, Mojtaba Seyedhosseini, Tolga Tasdizen, Lee Kamensky, Radim Burget, Vaclav Uher, Xiao Tan, Changming Sun, Tuan D. Pham, Erhan Bas, Mustafa G. Uzunbas, Albert Cardona, Johannes Schindelin, and H. Sebastian Seung. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in Neuroanatomy*, 9:142, nov 2015.
- [4] Funke. Jan, Stephan Saalfeld, Davi Bock, Srini Turaga, and Eric Perlman. CREMI: Circuit Reconstruction from Electron Microscopy Images, 2016.
- [5] Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, Jos?? Angel Conchello, Seymour Knowles-Barley, Dongil Lee, Amelio V??zquez-Reina, Verena Kaynig, Thouis Raymond Jones, Mike Roberts, Josh Lyskowski Morgan, Juan Carlos Tapia, H. Sebastian Seung, William Gray Roncal, Joshua Tzvi Vogelstein, Randal Burns, Daniel Lewis Sussman, Carey Eldin Priebe, Hanspeter Pfister, and Jeff William Lichtman. Saturated Reconstruction of a Volume of Neocortex. *Cell*, 162(3):648–661, 2015.
- [6] Jeffrey P Nguyen, Frederick B Shipley, Ashley N Linder, George S Plummer, Mochi Liu, Sagar U Setru, Joshua W Shaevitz, and Andrew M Leifer. Whole-



brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, (9):33, 2015.

- [7] J G White, E. Southgate, J N Thomson, and S. Brenner. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London*, 314(1165):1–340, 1986.