# MAPPING HISTOLOGICAL BRAIN IMAGES TO THE ALLEN MOUSE BRAIN ATLAS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jing Xiong
December 2019

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Mark Horowitz, Primary Adviser**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Liqun Luo**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**John Pauly**

Approved for the Stanford University Committee on Graduate Studies.

**Patricia J. Gumport, Vice Provost for Graduate Education**

*This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.*

# Abstract

In this thesis, we present an automatic framework that maps a histological image sequence to the Allen Mouse Brain Atlas which is in use at multiple neuroscience labs. The method is stable and continues to work on noisy sectional brain data. We also describe applications of our framework on multiple biological studies consisting different kind of datasets advancing research in understanding neurobiological organization. Code and sample dataset are available at our project website: `sites.google.com/view/brain-mapping`.

Histological brain slices imaged by high-resolution optical microscopy are widely used in neuroscience to study the anatomical organization of neural circuits. Systematic and accurate comparisons of anatomical data from multiple brains and from different studies can benefit tremendously from registering histological slices onto a common reference. To this end, the Allen Mouse and Human Brain Atlases have been created. However this registration task is extremely challenging due to heterogeneity of biological structures, image distortions and imaging artifacts introduced during the brain sectioning, staining, mounting, and imaging processes. Existing methods rely on an initial full or partial reconstruction of the experimental brain volume before registering a histological slice sequence to the reference, or extensive manual inspection is needed. Because these slices are often sectioned with non-standard angles, and without an external reference, curved structures end up straightened. Reconstruction is often inaccurate. Due to the low signal to noise ratio, traditional nonrigid image registration methods are not always reliable either.

In this thesis, we describe a framework that completely solves the z-shift problem and produces stable and accurate registration between histological image sequences and the Allen Mouse Brain Atlas. This work first determines the cutting angle and finds the best matching slice of each histological brain image in the reference volume directly by leveraging brain structural characteristics and symmetry. After finding the plane-wise mapping, we then register every image pair — each experimental slice and its corresponding cutting planes in the Allen Mouse Brain Atlas — nonrigidly to build a pixel-wise mapping. We modify the standard Markov random field framework on medical image registration to model accumulated tension when deforming tissue to more naturally deal with the easily-deformed cavities throughout the brain. We directly place control points on the automatically extracted salient points avoiding excessive deformation. Both steps novelly use L2

norm of histogram of oriented gradients difference as the similarity metric.

We have experimented our method on both simulated and experimental brains and on both sectional and full brains. Since our method is mostly automatic, and the accuracy is similar to or better than expert neuroscientists even for datasets where many slices are corrupted, our method is under use in different neuroscience labs of different institutions, making multi-brain histological data analysis possible and accurate, therefore advancing the research in understanding neurobiological organization.

# Acknowledgments

The past seven years at Stanford have been an unforgettable and invaluable experience to me. First and foremost, I want to express my greatest gratitude to my advisor, Mark Horowitz. I joined his lab during my second year and started my thesis project which I did not knew would take so long. I wanted to graduate sooner but now when I have reached a conclusion point, I wish the PhD time could be longer because I have started to miss Mark already. Like an old Chinese saying goes: give a man fish, and you feed him for a day; teach a man to fish, and you feed him for a lifetime. Mark has been a real mentor. He patiently guided me during my PhD and unreservedly taught me how to "fish".

I would like to thank my co-advisor Liqun Luo. Liqun identified this new problem in neuroanatomical research, and it later became the my thesis project. He gave me tremendous help on the biology side and initiated my collaboration with Jing Ren so that my tool can be used in a real application. He is very smart yet very diligent. Great minds are always similar. Working with him and Mark always motivates me to acquire new knowledge and to be open to new ideas and new fields.

I met Jing Ren a few years ago when I needed a neuroscientist to work closely with and to verify my results. It has been a great pleasure to work with her. She is a postdoc in Liqun's lab and will soon become a professor. She is very professional and always points out where I could improve. I learned teamwork through our collaboration. I feel very privileged to work with her, since her work has been the greatest promotion of my research. After her work is published, I got more collaborators and people who are interested in my work.

I would like to thank Professor Xiaoke Chen and Qian Wang. It has been a great pleasure to collaborate with them towards the end of my PhD. From the new collaboration, I got to understand my thesis project better and was able to further improve it. I also would like to thank Xiaoke for his time chairing my thesis defense.

Stanford is a very friendly community. Many professors have offered great advice. It has been a great pleasure to rotate with Professor John Pauly who also served as my reading committee and Professor Howard Zebker who is very nice and of great help. I also want to thank Professor Sandy Napel who served on my oral committee and gave advice on result evaluation.

I feel very grateful that people are interested in my project and are willing to use it. I thank my

fellow neuroscientists: Professor Hongwei Dong, Muye Zhu, Professor Alex Schier, Vassilis Bitsikas, William Joo, and Amelia Hidalgo.

Mark has been a great leader. It has been a pleasure to be in his group where all labmates are very supportive and easy-going. Every group outing has been an unforgettable memory, and it has always been interesting to discuss different topics and exchange ideas. I want to thank the whole VLSI group for being like a big family and keeping me from feeling lonely.

Outside of the VLSI group, I want to thank Yanwen Wu for being my best friend at Stanford, Tony Zhang for being a very supportive friend, and my recently deceased friend Yi Wan. It is very lucky to have friends in front of whom I could just be very vulnerable and freely talk about anything. Stanford is full of smart people. I could always learn something from everybody around me. I have made many good friends here who always give advice here and there and making me a better version of myself. In particular I learnt a lot from Qifeng Chen and Danqi Chen.

Lastly, I would like to thank my parents for giving me a lot of freedom and always being supportive about my decisions.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

The brain is the most fascinating organ in our body. In neuroanatomy, neuroscientists try to solve the magic of how we perceive things. Our brain is a very complicated system. There are about $10^{11}$ neurons in our brain [43]. In the world of neuroanatomy, neuroscientists are interested in where the neurons are, how are they connected to each other, and what do they do. Mouse brains are the most commonly used animal model in neuroscience, because of a mouse's fast reproduction, and its high-percentage similarity to the human genome sequence. We can start to understand human brain through understanding mouse brain. It is also a very complicated system consisting of about $10^8$ neurons [43]. To see inside an experimental mouse brain and perform neuroanatomical studies, neuroscientists conventionally section an experimental mouse brain into histological sections. With the help with different staining agents and a fluorescent microscope, neuroscientists can see the internal structures of a brain section as shown in Figure 1.1 as well as the neurons of interest. This is usually done by staining the same brain slice with different staining agents and image the same slice with different fluorescent channels. For example, the NISSL stain [23] stains every neuron's cell body and shows the structure of different brain regions; neurons expressing green fluorescent protein can be visualized with another fluorescent channel revealing individual neurons of interest.

It is crucial to standardize and digitize anatomical information to allow data from multiple brains to be compared in the same reference brain. To this end, detailed anatomical brain reference atlases have been established for both human and animal model studies [8, 28, 38, 59]. Given these references, imaging and digitization are usually followed by annotating data based on histological stains in comparison with a brain reference, for example, the Allen Mouse Brain Atlas [2, 38].

Ideally, all experimental brain images would be automatically registered to an anatomical reference volume, creating a platform for the comparison and integration of results from different experiments. Existing programs mapping a 2D histological sequence to a reference volume often require an initial reconstruction from these partially corrupted slices and therefore only work well with datasets of very good quality. Problems with automatic methods increase when an experimental brain is sectioned

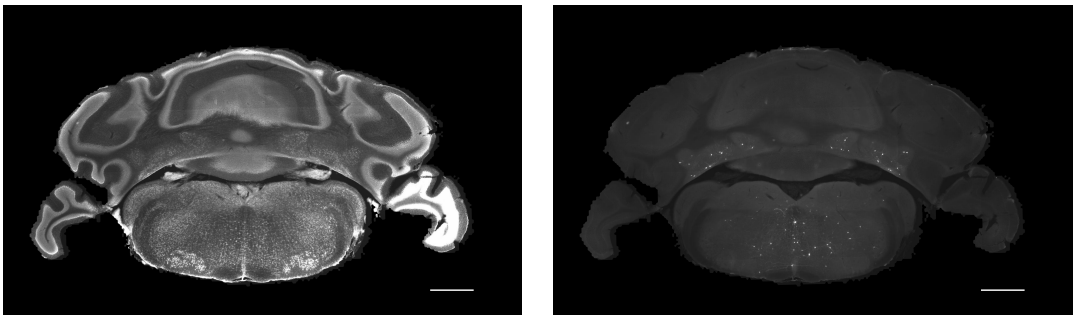


Figure 1.1: Example experimental slice with imaged with two different fluorescent channels. Left: Nissl-stained image stains every cell body of neurons and the structure of brain regions. Right: Same slice imaged with green spectrum shows neurons expressing green fluorescent protein; bright spots are individual neurons of interest. Scale, 1 mm.

with a non-standard angle or when brain histological data-sets suffer from common slicing artifacts. However, histological datasets often require months of experiments to generate, so all the data is precious. As a result, most labs still rely on manual brain region identification to fully utilize all of the experimental slices even if they are partially corrupted. This often requires a neuroscientist to go through a sequence of slices in the reference until they can confidently locate a neuron's region back into reference and often requires some 3D transformation in mind if the sectioning angle is different than the standard reference brain.

In order to understand the neural connection, neuroscientists have developed different neural tracing techniques. In recent years, there have been breakthroughs in viral genetic mediated tracing. Figure 1.2 shows a horizontal view of a mouse brain in which you can see presynaptic partners of the dorsal raphe neurons at a whole brain level. These new techniques can help neuroscientists understand the brain at a whole brain level, but also present severe challenges for them to analyze their results. If a mouse brain is sectioned into 50um-thick coronal slices, one will get more than 200 slices. In addition, multiple animals are needed to gather enough results for a robust result, which creates a large amount of data to be compared to the reference and localized. If this comparison is done manually, each result requires a very time consuming analysis for the neuroscientist. In fact, in Weissbourd *et al.* 's study [64] where Figure 1.2 was generated, the authors reported manually localizing all the neurons of interest. Manual localization is labor intensive, time consuming, variable, and sometimes subjective. In addition, it is also irreproducible — even the same person at different time may place a neuron at a different location than the last time, let alone other labs trying to reproduce the results.

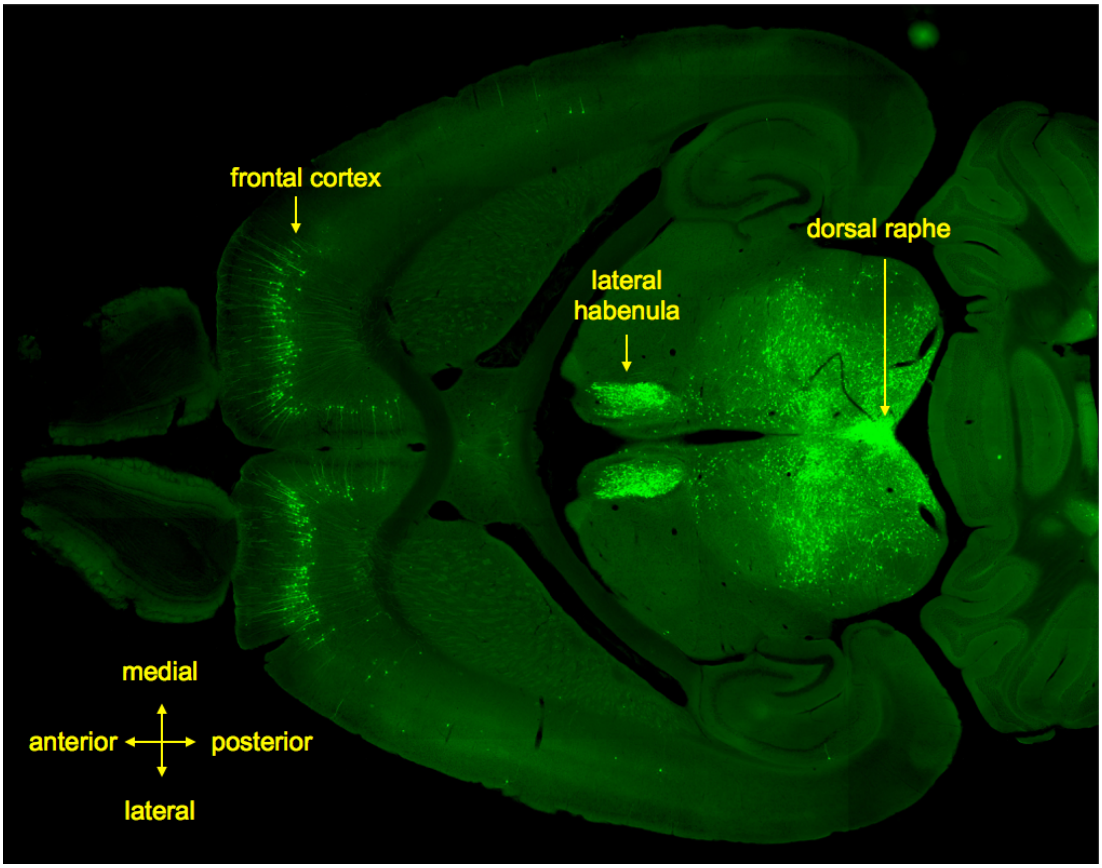


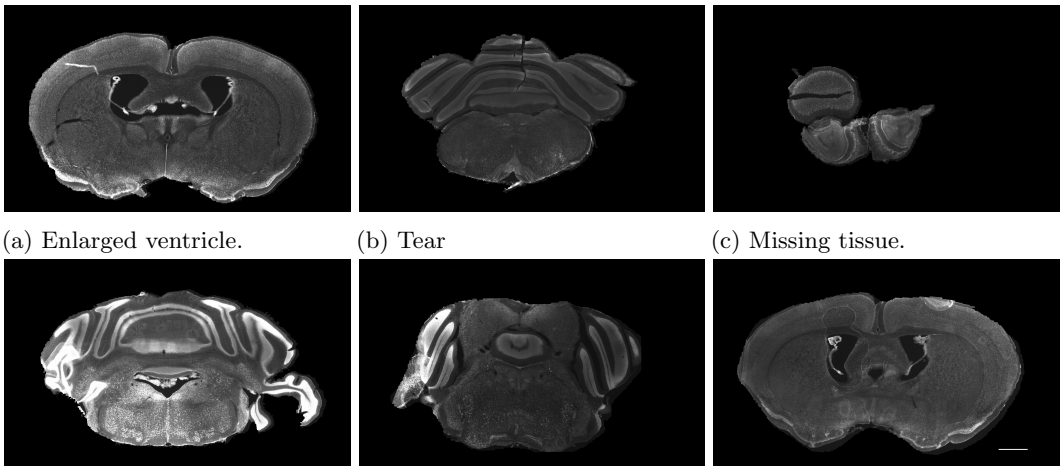

Figure 1.2: Presynaptic partners of dorsal raphe serotonin neurons. Long-range input from frontal cortex and lateral habenula to dorsal raphe serotonin neurons revealed by rabies-mediated trans-synaptic tracing in this horizontal section of the mouse brain. [3]

## 1.1 Problem Formulation

The lack of a program that can robustly map experimental slices to a brain reference has become a bottleneck in processing data from multiple brains. In this thesis, I will describe our method to remove this bottleneck: a program that automatically maps a sequence of coronal experimental images of mouse brain to the Allen Mouse Brain Atlas. Our approach handles imperfections common in histological slice data.

**Experimental Dataset** The experimental datasets that we work with are generated with the conventional sectioning method. An experimental mouse brain is first extracted from the mouse skull after some perfusion and dehydration processes which, depending on the processing time, may enlarge (shown in Figure 1.3(a)) or shrink empty spaces in the brain. The brain is then mounted into a plastic box with mounting solution. After freezing, the frozen brain is taken out from the mounting box and fixed on the cutting platform and cut coronally into sequential sections with a uniform thickness which is typically between 50 µm to 60 µm. Depending on how the brain is mounted and placed on the cutting platform, there is usually a discrepancy between the real cutting angle and the standard coronal section cutting angle. The sections are then placed into a washing liquid so that the mounting liquid dissolves. The real tissues are then transferred onto a slide holder with a brush. These cutting, dissolving, and transfer processes often introduce artifacts — tears (Figure 1.3(b)), misplaced tissue and folds (Figure 1.3(d)), and tissue distortion — into our experimental images. Nissl stain is then applied. Depending on the staining time and how evenly the stain is applied, tissues may be unevenly stained (Figure 1.3(e)). Finally a coverslip is laid on top of the slide holder, which sometimes results in air bubbles as shown in Figure 1.3(f). The tissues are then imaged with a light microscope at a resolution of around 5 µm per pixel. Given the mouse brain size and the thickness of our coronal slices, sectioning a full brain usually results in around 200 images with dimensions around 1000 by 1000 to 2000 by 2000 pixels.

**Allen Mouse Brain Atlas** Similar to our experimental datasets, the Allen Reference Atlas [2, 38] (referred as the Allen Brain Atlas or ABA in this thesis) was created with standard procedures of isolating, cutting, fixing, and pretreating. The orientation of the brain was adjusted using a dissecting scope, and sectioned at 25 µm. Slices that passed quality control are kept for use [1]. The Allen Atlas consists of three volumetric data files: the atlas volume — a grayscale Nissl volume of reconstructed brain at 25 µm resolution, the annotation volume — a structural annotation volume at 25 µm resolution, and the grid annotation volume — a structural annotation volume at grid 200 µm resolution for gene expression analysis (not used for our thesis). The atlas volume or the Nissl volume was reconstructed by aligning high-resolution coronal Nissl images to each other and aligning the reconstructed volume to a lower-resolution MRI scan of a mouse brain. This first-stage reconstructed volume was then aligned with a sagittally sectioned specimen. After a straight mid-sagittal plane



(a) Enlarged ventricle. (b) Tear (c) Missing tissue.

(d) Misplaced tissue and fold. (e) Uneven staining. (f) Air bubble.

Figure 1.3: Artifacts introduced during the image generation processes. Scale, 1 mm.

was achieved, a synthetic symmetric space was created by reflecting one hemisphere to the other side of the volume. The annotation volume is generated by extracting structures from 2D coronal reference atlas plates and interpolating to create symmetric 3D annotations. This Allen Atlas was further improved or smoothed after the introduction of a common coordinate framework in 2015, which is a population average that supports the integration of new mouse brain datasets in the Allen Brain Atlas Portal by deformably aligning the Nissl volume and the annotation of the existing Allen Reference Atlas to the new common coordinate framework. [2] This thesis uses the 2015 version of the Allen Atlas. The data is downloadable at `help.brain-map.org/display/mousebrain/API` which also includes processing instructions. Despite the detailed models that Allen Institute provides, they do not provide a computational tool to help neuroscientists map their experimental data to the Allen Atlas automatically.

Now we have described the generation process of the experimental dataset, we can be more precise about our goal — to build an automatic program that maps a sequence of distorted coronal histological sections of mouse brain which has high-resolution in the sectioning plane (determined by the resolution of the imaging system) and comparatively low-resolution along the sectioning axis (limited by the minimum slice thickness) to the grayscale Nissl volume of the Allen Mouse Brain Atlas. The thesis describes both our method and applications of our method to real experimental datasets.

## 1.2 Thesis Outline

In Chapter 2, we first give an overview of the existing methods in biomedical image registration, then specifically mouse brain histological image registration. The drawbacks of these methods then lead to our goal definition. Chapter 3 describes our method in detail. In Chapter 3, we first discuss image characteristics of the histological mouse brain image, then describe our method's basic building blocks — 2D to 3D plane-wise mapping, then the 2D nonrigid pixel-wise mapping, and finally the improved 2D nonrigid registration. This chapter also includes performance of our method on simulated datasets. In Chapter 4, we discuss how we can apply our method to different kind of datasets from different labs of different institutions and covering different brain regions. We also describe how our methods can be extended to specific properties of some real datasets. This chapter also includes evaluation of our methods on real experimental data. Both chapters are partially based on our published work [66]. In Chapter 5, we discuss some of the issues we have found during our experiments on different datasets, as well as other directions that we have tried that were not as successful. The last chapter concludes our work. In the appendix, we include some links to the Allen Brain Atlas resources and how we used them in our work.

# Chapter 2

# Existing Image Registration Approaches

## 2.1 Biomedical Image Registration

Registration is a process that establishes a spatial correspondence between two images so that both can be viewed in the same spatial frame. Usually one image is designated as the reference image or the fixed image. A transformation is applied on the other image — called the moving image — so that it is aligned with the reference image. In biomedical applications, it is necessary before you can make a quantitative comparison between two scans or images. Such mappings are essential within diagnostic settings, treatment procedures [44], functional studies, and in biomedical research [63]. Traditional registration framework usually involves iteratively comparing the fixed image with the original or warped moving image with a similarity metric, evaluating the cost function and derivatives, optimizing to update the transform parameters, applying the new transform, and interpolating new intensity values, as shown in Figure 2.1.

Commonly used similarity metrics can be divided into two groups — intensity-based and feature-based. Intensity-based methods compare images based on their intensity information. Based on if the fixed and moving images are of the same modality, or equivalently acquired by the same scanner or sensor type, the choice of an image similarity metric would be different. Commonly used similarity metrics for same modality images are sum of squared intensity difference and cross correlation. For multimodal image registration, the most common metric is mutual information or normalized mutual information. Intensity-based methods register images as a whole. In contrast, feature-based methods first find corresponding image features, such as points, lines, curves, and patches, and then transform the moving image to bring the corresponding features together. These corresponding features are often computed with well-established feature descriptors, such as scale-invariant feature transform

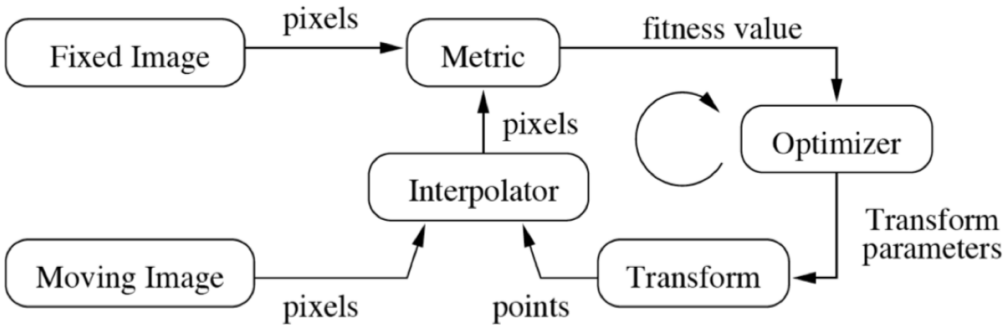



Figure 2.1: Registration flowchart. [30]

(SIFT) [42] and maximally stable extremal regions [46].

Based on the deformations possible, different levels of transformation are computed, starting from the simplest rigid transformation, where images are only rotated and translated, to the non-linear transformation — elastic or nonrigid, where the transformation can distort an image locally to match with another. When a deformable registration is needed, a linear transformation — usually affine transformation which includes rotation, scaling, sheer, and translation — is computed to align the images globally, and then the nonrigid registration is computed to align the images locally.

As machine learning has become more popular, biomedical fields have seen more and more use mostly in the segmentation and detection tasks [41]. The registration applications are split into finding a similarity measure [65, 57] or transformation parameters [49, 67]. There have also been research on using deep learning to compute nonrigid registration [39, 12, 40], but none of this work has been a breakthrough in applying deep learning to biomedical image registration. These modest results might be the result of the vast variety of imaging modalities and the limited available data.

## 2.2 Histological Image Registration

Different from the regular image registration problem, what makes this problem a unique and a challenging research topic is that the moving image is not readily prepared to directly register to the fixed image. The moving images are a sequence of physically sectioned 2-dimensional images from the mouse brain, and the fixed image is an isotropic volume of a reference mouse brain. Therefore this registration problem is between a sequence of distorted 2-dimensional images and a 3-dimensional volume.

The problem of mapping a sequence of histological slices to a reference brain has been well studied [52]. This prior work first reconstructs an initial volume estimate from the slices and then

registers this reconstructed volume to the reference, because this is the most straightforward approach [19, 48, 58]. Other work discusses the reconstruction problem directly in the absence of reference volumes [7, 16, 32, 50]. The challenging problem now is to create this initial volume estimate because registration between a reconstructed volume and a reference is relatively standard. Initial work reconstructed the experimental volume by pairwise registration of adjacent slices [16, 50, 58]. Due to tissue distortions, rigid registration is not sufficient. But pairwise nonrigid registration propagates any registration errors throughout the whole brain. This is especially problematic if any slice has a large deformation such as missing tissue.

To improve overall reconstruction results and reduce error propagation, some methods align each slice with multiple neighboring images. For example, Ju *et al.* [32] reduced error propagation by warping each slice with a weighted linear and nonlinear combination of warp fields to multiple adjacent slices. Others used blockface images [19] or selected internal reference slices to reconstruct small chunks and then put together the entire volume [7, 48]. However, with almost every slice at least slightly distorted, internal nonrigid registration will likely change the original shape of biological structures. Because this process tries to maximize the similarity between adjacent thin, *e.g.*40 µm to 60 µm, histological slices, curved 3D structures along the sectioning direction may end up straightened. This 3D structure-straightening problem is known as the banana problem or z-shift [6, 45]. Once this error is introduced, it is hard to reverse completely even when this volume is registered to the reference. To avoid these volume distortion errors, one needs to use the reference volume earlier in the process by registering each experimental slice to its corresponding sectioning plane in the reference volume.

Now the main challenge is finding the corresponding plane for each slice [68]. This task is made more difficult because the experimental volumes have a non-standard sectioning angle, the brains are tilted in the sectioning machine and have an anisotropic resolution. The reconstructed volume has a very high resolution in the sectioning plane (determined by the resolution of the imaging system) and comparatively low resolution along the sectioning axis (limited by the minimum slice thickness). The slice-to-slice approaches usually assume that cutting planes are parallel to the acquisition planes of the 3D medical image [52] or at least the cutting angle of the microtome are constant through the cutting process [22, 25, 68]. Abdelmoula *et al.* [4] used one of the most prominent features — the hippocampus — to determine the best match plane for each experimental tissue section, however, the search space is limited, and the cutting angle difference is not considered. Papp *et al.* [51] developed an interactive tool with which neuroscientist can reslice a reference volume with adjustable angle and position. Five to ten slices are mapped, then the remaining slices are interpolated. Other recent work used an iterative approach by first reconstructing a small volume and registering these slices to their corresponding planes in the reference brain [25, 68]. Yang *et al.* [68] selected a reference slice that maximizes the normalized mutual information after a 2D rigid registration is performed between a histological slice and each MRI slice. Goubran *et al.* [25] registered each histological slice of a

human brain to its corresponding MRI slice after blocks of the human brain were registered. Possum [45] developed an open source software framework that reconstructs a volume with or without using an external reference. While these methods work well when the sectioning angle difference is small, they introduce errors at larger sectioning angles.

No matter if the first step is to reconstruct the volume or is to find corresponding plane for each histological image, the next step is registration. Free Form Deformation (FFD) [54, 55] has been the most common deformation or transformation model in neuroscience studies to map histological brain images [4, 5, 17, 20, 21, 31, 62]. Mutual information is often used as the similarity metric to register histological slices because of image appearance difference caused by acquisition procedure variability. This method is highly dependent on the initial condition because the choice of using mutual information as a similarity metric often leads to a highly non-convex optimization problem with many local minima [26]. Because of staining variability within a slice, using mutual information does not always work.

# Chapter 3

# Mapping Histological Images without 3D Reconstruction

To avoid the issues described in Section 2.2, we concurrently estimate the sectioning angle difference and the best matching planes in the atlas volume for each slice. This approach requires us to find the best matching slice in the reference before applying nonrigid deformations. Since the resulting slice comparisons are noisy, we aggregate information from all slices and use information about the brain's structure to find the best match. Our method does not have a reconstruction step, therefore completely eliminating the z-shift problem.

After each matching reference slice has been determined, we need to perform a 2D registration between it and its matching histological slice. Instead of using mutual information, we find that the L2 norm of histogram of oriented gradients (HOG) [18] difference suits histological slice properties better. Because HOG is non-differentiable, we base our work on the elegant discrete Markov random field (MRF) approach in [24]. Building on the tissue labeling information of the annotation volume of Allen Brain Atlas, we build a MRF model based on tissue coherency. To make the warp field even more realistic, we place control points only on significant features that are consistent in both our histological images and the reference atlas image.

Our strategy makes the maximum use of the reference volume, successfully deals with the non-standard sectioning angle problem, preserves the curvature of the object — eliminating the z-shift problem [6], and is more tolerant to data corruption. This method takes into account some of the brain's structural properties to minimize error, including the compressibility of different brain regions. The algorithm is tested both on the full brain and sectional brain data, yielding faster and better correspondence than possible before.

This chapter describes in more detail how we choose the similarity metric, find the sectioning angle difference and the best matching plane in the reference volume for each histological slice, and

nonrigidly register each slice to the corresponding sectioning plane in the atlas. First we describe the similarity metric that we use.

## 3.1 Histological Slice Similarity

Both in the 2D to 3D localization and the 2D nonrigid registration steps, a similarity measure is needed. We first take a look at different commonly used metrics in biomedical image registration. The first is sum of squared differences. It is used to register images acquired by with the same modality. Even though our images are Nissl-stained, the staining time and imaging settings are different, therefore the intensity of a same biological feature does not correspond. Since when finding the best corresponding slice, we compare an experimental slice to a sequence of potential matching atlas slices, even if the images are not exact of the same modality, we may still be able to find the best slice as long as the slice generates a smaller difference value than other candidates. Another commonly used intensity-based metric is the correlation coefficient. The correlation coefficient of real intensity image A and image B is given as:

$$\rho(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}(\frac{A_i-\mu_A}{\sigma_A})(\frac{B_i-\mu_B}{\sigma_B}) \tag{3.1}$$

where $A_i$ and $B_i$ are intensity values at the same pixel i, $\mu_A$ and $\mu_B$ are mean intensities of image A and B, and $\sigma_A$ and $\sigma_B$ are standard deviations. Since brighter and darker features tend to be consistent in the Nissl slices from the two volumes, correlation coefficient which measures the linear dependency of intensity values from two images may also work. Since our experimental images and atlas images are not acquired from single modality, we also looked at the most commonly used metric for multi-modal images — mutual information or normalized mutual information. Mutual information measures the mutual dependence between two random variables. The state of art to register histological images is to use normalized mutual information [20, 31]. Unlike the correlation coefficient, instead only looking at the mean and standard deviation, intensity values are first binned into a user-defined number of bins to estimate their probability distribution for calculation of the marginal and joint distribution and further the mutual information.

We evaluated these metrics' potential on real experimental images. To estimate the "ground truth" for this dataset, we rotated the atlas with different rotation angles around both the superior-interior axis and the horizontal axis, resectioned the rotated atlas into 25 µm slices, and reindexed them. Taking a full experimental brain with 220 50 µm-thickness slices, we selected more than 10 slices covering the full brain while avoiding the slices with significant artifacts and asked a neuroscientist to identify first the rotation angles that match the cutting angles of the experimental volume. Once the selected the cutting angle, they gave a range of reindexed slices that may potentially be the corresponding slices of the selected experimental slices. A range of roughly 5 to 10 slices were selected

for each experimental slice because adjacent atlas slices are very similar, and even a neuroscientist cannot choose one slice to be the best corresponding slice for a given experimental slice.

With this estimate of the matching slices,to check a metric's potential, we enlarged the range to 60 slices and computed difference between each experimental slice to its potential matching slices with each metric. The results were not encouraging. With sum of squared differences and correlation coefficient, the slice that gives minimum difference or maximum coefficient tends to be the one that matches the overall shape of the experimental slice; internal structure matching was not a significant factor. That is lower score slices often looked qualitatively better. Therefore the two metrics are not trustworthy, since distortions greatly affects the results. Despite mutual information's wide use in registering histological brain slices, it does not work well in our experimental images because this metric fails when intra-slice uneven staining creates intensity variability within a structure, which breaks the statistical correlation between a slice and its target image.

From these experiments with intensity-based metrics, we learned that we need to deal with the distortions and also the uneven illumination. If we continued using intensity-based methods, we would need to first correct the distortions; otherwise distortions would likely dominate the difference score but not the real structural variation. But correcting distortions brings two other problems. First, correcting distortions involves nonrigid registration whose computation time is huge. Second, without a reference slice we would not know if we are correcting distortions only or making two slices more similar than they should be. Therefore we decided to see if there is a metric that can measure similarity even with the existence of distortion. This new approach also needs to deal with the uneven illumination which we find hard to remove because of its coexistence with biological structures of different intrinsic illumination.

To address these issues, we stopped looking at more intensity-based metrics and started to try feature-based metrics. Many descriptors are designed to be somewhat distortion-tolerant, not as sensitive to uneven illumination. The first one we tried is the scale-invariant feature transform (SIFT) descriptor [42]. While this approach works for many natural images, it did not work on our images. Many matching points the metric detected are not visually correspondent at all. One reason is because SIFT is scale-invariant and rotation-invariant. We do not need these properties, because with the global affine registration, we can resize a image and rotate the image upright. This extra flexibility, combined with the higher noise levels in our images worked against SIFT features. Making the problem worse was the fact that the atlas volume was created by putting together coronal sections sectioned with a similar manner. Even though the volume is postprocessed to be much smoother. After rotation, imperfectly aligned slices result in jaggedness which further degrades results. SIFT is too good for our images.

From these experiments, we discovered characteristics of our images that were visually insignificant, and need to be tolerated by our feature metric:

1. Staining reagent and microscopic setting difference can cause direct comparison of intensities

to be not useful. Even worse, due to the non-uniformly applied staining reagent, some slices are unevenly illuminated.

2. Nissl-stained [23] images only highlight the cell body of neurons. Two matching images will show corresponding anatomical structures but do not have pixel-wise cell body level correspondence.

3. Sparsely scattered or densely populated cell bodies make images low-contrast and noisy. Many descriptors that work with man-made scenes do not perform well.

4. Distortions caused by brain's elasticity require metrics that work even when the two images are slightly distorted from each other. This distortion tolerance also allows it to compare a distorted histological slice to a reference slice.

5. Even though the newest Nissl volume of the Allen Mouse Brain Atlas (2015) is smoother than the Allen Mouse Brain Atlas (2011), it is still Nissl-stained volume constructed from physically-sectioned mouse brain slices and is not perfectly aligned. So an ideal metric should be somewhat tolerant to this imperfect alignment.

To conclude, we need a feature-based metric that is somewhat tolerant to distortion, not scale-invariant, and not rotation-invariant. This metric does not measure directly absolute intensity values, but needs to generally describe the relative contrast and gradient orientation information to be robust to the uneven illumination and noise. Therefore we continued looking at feature-based descriptors that were invented before SIFT, and Histogram of Oriented Gradients (HOG) [18] came to our attention. This descriptor is widely used in human detection and has capability to deal with pose, illumination, and background variations which mimic many of the issues in our images. The image patch to be described is divided into small cells. Each cell is described with a histogram. To form this histogram, each pixel's magnitude of gradient is then binned into different bins according to its gradient direction. HOG describes image patches in blocks. The block size is defined by the number of cells in the horizontal and vertical direction. To generate description of a block, each histogram is normalized based on the magnitude of the histograms of its neighboring cells in the block and then concatenated. HOG describes a small patch rather than individual pixels. Gradient binning gives some flexibility to distortions but still captures the overall direction of edges. To make our metric more distortion-tolerant, we use a relatively large cell size. Normalizing a cell's magnitude by the magnitude of its neighboring cells reduces the negative effect from uneven staining. It also well accommodates the unsmoothness nature of the atlas volume. Because even if an unsmooth volume is rotated, even though edges are not perfectly aligned, true gradients are still kept.

**Affine Registration** To use the HOG metric, the two images first are brought to the same coordinates with an affine transformation. Smooth tissue contours are extracted by applying the Fourier transform on the boundary curve and removing high-frequency components. We then find

contour point correspondence with Shape Context [9]. Shape Context first extracts randomly a set of boundary points with oriented tangents on both images. Each point is then described by a histogram of the relative position of this point in terms of all other sampled points in a log-polar space. Point matches are found by minimizing the total cost between two sets of points where the cost between two points are measured by the chi-squared test using the Hungarian method [36]. With the point correspondence, we can estimate an affine transformation between two images. Since sheering is not desired, our modeling of the affine transformation is different from that in the original Shape Context paper. We first find a similarity transformation (including rotation, uniform scaling, and translation). Scaling is estimated as the ratio of mean distances between points within each point set. Rotation and translation are estimated with the Umeyama method [60]. Now the experimental images are straight up. To accommodate the global deformation caused by the force in the direction of sectioning, slices are further rescaled in horizontal and vertical directions so that the height and width of the two slices are the same.

After globally aligning two slices, we use HOG to describe them with the same cell size and same block size. The difference between the two slices are then measured by the Euclidean distance between these two HOG representations. For 2D to 3D localization, we use a large cell size so that the metric is less sensitive to local distortions. Our experiments show using HOG difference, the best matching slices are within or close to the index range selected by the neuroscientist for most slices. Of course optimizing the best matching slices as a whole is needed to further improve the results.

## 3.2 2D to 3D Localization

Since histological slices are often cut with near constant angles with a microtome [22, 25, 68], it is fair to assume a constant cutting angle throughout the whole brain. This helps us reduce the search space. Because the atlas is uniform in each dimension, to find the cutting angle difference, we rotate the atlas with different angles, resection it into coronal slices, reindex the slices in order, and compare the new resectioned atlas slices to the histological sequence. The following sections give our dynamic programming formulation to solve the alignment problem to determine the slicing angle, and a simple method to increase sensitivity to angular shifts.

### 3.2.1 Slice Mapping with Dynamic Programming

The best cutting angle is the angle that maximizes the similarity between all histological slices and their corresponding best matching slices in the atlas. Because in-plane rotation are handled by our slice alignment method, we only consider rotation angle $\alpha$ about the superior-interior (y) axis and $\beta$ about the left-right (x) axis. To solve the problem, we first find the best matching slice for each experimental slice given a potential cutting angle.

The problem can be represented as follows: Let $\mathrm{I}_{1\ \ldots\ \mathrm{N}}$ with spacing $\mathrm{s_E}$ be the experimental slice

sequence, and $V_A$ be an isotropic atlas with voxel dimension $\mathrm{s_A}$, defined on the domain $\Omega$. After rotating the atlas with potential best rotation $R_{\alpha\beta}$, we reslice the rotated atlas into coronal slices and re-index them as an atlas slice sequence $\mathrm{J}_{1\ldots\mathrm{M}}$. Using the L2 norm of HOG differences described in Section 3.1, we aim to find a mapping that matches each slice in I to a slice in J which minimizes the overall difference.

Taking into account potential compression along the longitudinal axis, slice quality variation, and intersubject variation, we formulate the problem with a single subset A of all slices I, where A is an ordered selection of $1 \ldots N$, which may or may not be the whole sequence of experimental slices (depending on the image sequence quality and the value of $\mathrm{s_A}$ and $\mathrm{s_E}$). A is chosen to span the full sequence while avoiding damaged slices. Because the experimental sequence is ordered, as is our atlas sequence, and the experimental volume and the atlas volume represents both full mouse brain, for each experimental image we can roughly estimate its best matching slice could be in the atlas volume as long as the cutting angle is not way different. This defines our searching range in terms of new indices in the rotated and resectioned atlas for the experimental images. Of course, these are substantial searching ranges, and there may be some overlap of searching ranges between nearby slices. We compute the HOG difference between each experimental slice in A and its all possible best matching slices given by the searching ranges and get a matrix of HOG difference scores where each row and each column represents experimental slice and atlas slice respectively. To find the best corresponding positions of experimental slices is equivalent to finding the index of resectioned atlas images for each experimental image index in A that minimizes the summation of HOG difference scores.

Since the experimental image stack is ordered, the best corresponding slice of an experimental slice can only be after the best corresponding slice of the experimental slice before it. We can compute distance between two adjacent experimental slices selected in A with their indices and the slice thickness. Since both volumes represents adult mouse brains, based on this known distance between two experimental slices, we can define the minimum and maximum possible distances between the corresponding slices of these two experimental slices in the atlas. Therefore we can further restrict the searching range after the corresponding slice of the previous experimental slice is determined. With these properties, we formulate this slice mapping and difference minimization problem as a dynamic program.

Let $\mathrm{I_A}$ be the ordered selection of experimental slices, and let J be the resliced atlas sequence ordered from the same direction along the longitudinal axis and spacing $\mathrm{s_A}$. The cost, $C(i,j)$, is defined as the minimum cost of mapping the first i slices in A to a sequence of j slices, where the $i^{th}$ slice has to be mapped to the $j^{th}$ slice:

$$C(i,j) = \begin{cases} \rho(I_{A_i}, J_j), & \text{if } i = 1 \\ \min_k(C(i-1,k) + \rho(I_{A_i}, J_j)), & \text{else} \end{cases} \tag{3.2}$$

where $i \in A$, $0 \leq j \leq \mathbf{card}(J)$, $\rho(a, b)$ denotes the difference score between Slice a and Slice b measured with the HOG similarity metric. There is no spatial constraint on the first slice. Because we are not interested in finding the slice that gives minimum difference, and we are interested in finding the slices that satisfy the spatial constraint and gives the minimum difference sum, we directly record the cost of matching the first experimental slice to every possible slice in its searching range. We do not make any decision at this time. We continue to look at the next slice and solve the problem of finding the best positions of the first two slices to minimize the HOG difference. Now when calculating the cost of matching two slices where the second slice has to be the $j^{th}$ slice, it already contains information of where the first slice has to be matched given by the distance constraint we have previously discussed. Formally this constraint is given by:

$$|\frac{s_{\mathrm{A}} * (j - k)}{s_{\mathrm{E}} * (A_{\mathrm{i}} - A_{\mathrm{i\text{-}1}})} - 1| < \theta \tag{3.3}$$

where $A_{\mathrm{i}}$ is the original index of the $i^{th}$ slice in the selected sequence and $\theta$ is a user-defined threshold value. This spatial constraint constrains the ratio of the distance between slices in the atlas and the experimental slices match to $\theta$. We repeat the process and fill in the cost for every possible pair of $i$ and $j$ until the last slice in A.

We denote the best $k$ that satisfies Eq. 3.3 and is used to fill in the cost matrix (Eq. 3.2) as $k^*$. The best intermediate steps are saved by updating the three-dimensional array $\mathcal{M}$ for each $i, j$:

$$\boldsymbol{\mathcal{M}}_{\alpha\beta}(i, j) = \begin{cases} \begin{bmatrix} j \end{bmatrix}, & \text{if} \quad i = 1 \\ \begin{bmatrix} \boldsymbol{\mathcal{M}}_{\alpha\beta}(i-1, k^*) & j \end{bmatrix}, & \text{else} \end{cases} \tag{3.4}$$

$\boldsymbol{\mathcal{M}}_{\alpha\beta}(i, j)$ lists the the indices in J that best match each of the first i slices in A, where Slice i in A is mapped to Slice j in J and the atlas is rotated with angle $\alpha$ about the y axis and $\beta$ about the x axis. The optimal mapping is therefore given by:

$$\boldsymbol{\mathcal{M}}^*_{\alpha,\beta}(I_A) = \boldsymbol{\mathcal{M}}_{\alpha,\beta}(\mathbf{card}(A), j^*) \tag{3.5}$$

where

$$j^* = \arg\min_j C(\mathbf{card}(A), j) \tag{3.6}$$

The cost of mapping all slices in A to resectioned slices in J with atlas rotated by $\alpha\beta$ is given by $C(\mathbf{card}(A), j^*)$.

With this final formulation we could generate slice matches that were all within the matching range picked by our neuroscientist. Prior to this final formulation, we also computed the best matching slice without the additional constraint, but the result was far from satisfactory. This poor result proves HOG difference, even though can handle distortion, alone is not very sensitive. The spatial constraint

allows the tool to aggregate many comparisons, which increases the effective sensitivity. Interestingly, when we tried to increase the spatial constraints, by adding a penalty proportional to the difference in the distance between previous experimental slice and the current experimental slice and the ideal distance between the previous best matching slice and the current best matching slice the results were worse. Our guess is that some distance in the experimental volume does not represent the same distance in the atlas. Because our HOG difference is not very significant, adding this penalty may overwhelm the real signal. There may be stretch and compression along the anterior-posterior axis as well. In addition, the slices are soaked in a resolving solution before being made into slides. When the slices are extracted, the front and back orientation are not easily identifiable because the tissues are transparent and nearly left-right symmetric which means that some of the slices can be flipped relative to other slices. These flipped slices are not easily fixed in post processing and also make slice matching harder.

### 3.2.2 Cutting Angle Difference Determination

After running this dynamic program with different sectioning angles we should be able to directly choose the angle that gives minimum cost score to be the best cutting angle. However, since HOG is relatively insensitive to local distortions, and each slice is slightly distorted, when summing up all the costs we also sum up a lot of noise. Therefore when the angle is very close to the true sectioning angle, the difference among neighboring angles is not substantial. To improve our robustness, we use a different approach. This approach also predicts how we should adjust the rotation and prevents exhaustive searching in the previous approach.

Biological structures change quickly along the posterior-anterior direction. It is not hard to tell if an experimental brain is sectioned with a different angle than the atlas volume, even if the angle deviation is only several degrees, because structures that appear in the same slice in the atlas will be in different slices in the experimental slice sequence. For example, if the left side of the brain is tilted to be more anterior, on average the right hand half coronal brain slice will appear to be more posterior to the left half. Thus if we match the left and right half slices of an experimental brain separately to the atlas, we will see that the slice number of the matching slices of the left half brain will be on average higher than that of the right half. To experiment, we took an experiment brain which is obviously not sectioned in the same angle as the atlas and tried to find their matching half slices in the unrotated atlas. We observed that the experimental slices have near left-right symmetry. Therefore we tried to find matching upper and lower half slices in the atlas.

In Figure 3.1, we show two column of slices. The left column shows (A) Slice 15 in the experimental volume, (B) Slice 55 in the atlas with a similar upper half to (A), and (C) Slice 85 in the atlas with a similar lower half to (A). The coronal atlas slices are ordered from posterior to anterior. The index of (B) is smaller than that of (C), indicating the upper half of the experimental volume is tilted more posterior than the atlas. Looking at (D) Slice 85 in the middle of the experimental volume

showing in the right column, the upper half slice looks like the upper half slice of (E) Slice 224 in the atlas, while the lower half looks like that in (F) Slice 257 in the atlas. The index numbers are bigger, but the difference between the numbers is similar to that of the left column indicating there is a consistent rotation in the brain. Based on this observation, we use matching slice index differences of half brains to determine if a rotation angle best fixes the cutting angle difference between the experimental brain and the atlas.

Because mouse brains have left-right symmetry, the rotation angle $\alpha$ about the superior-interior (y) axis tends to be around zero. The rotation angle $\beta$ about the left-right (x) axis tends to be larger because the mouse brain is not flat at the bottom and can easily be set tilted on the microtome plate. Here we use the determination of angle $\beta$ as an example; the flowchart is shown in Figure 3.2.

To find the best rotation angle $\beta$ about the x axis, we solve the slice mapping problem with the method described in Section 3.2.1 on the upper half slices and the lower half slices respectively. We take the index difference between the optimal mapping given by:

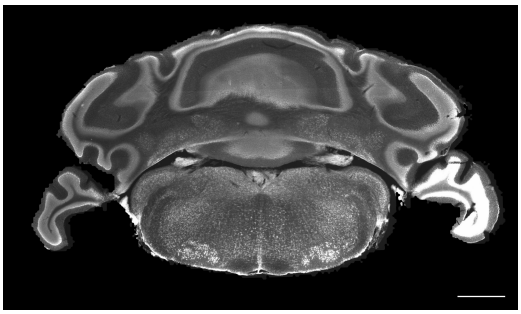
$$D = \frac{1}{\mathbf{card}(A)} \sum (\boldsymbol{\mathcal{M}}^*_\beta(M_{upper}(I_A)) - \boldsymbol{\mathcal{M}}^*_\beta(M_{lower}(I_A))) \tag{3.7}$$

where $\mathrm{M_{upper}}$ and $\mathrm{M_{lower}}$ denote binary masks to apply to both experimental image and resliced atlas image to include only half of a slice. Positive $D$ means upper half experimental slices are matched to slices more anterior than lower half slices. Therefore the atlas should be rotated more in the positive direction about the left-right (x) axis, where the positive direction is defined by the right-hand rule around the x axis. If $D$ is negative, then the atlas should be rotated in the negative direction. We reslice the atlas again after rotation and repeat the above steps until the index difference flips signs meaning we need to rotate the atlas in another direction. The rotating angle changes in a step size of one degree. When the flipping of sign occurs, we choose the angle between the current angle or the previous angle which gives the smaller absolute index difference. The same steps are repeated to determine $\alpha$.

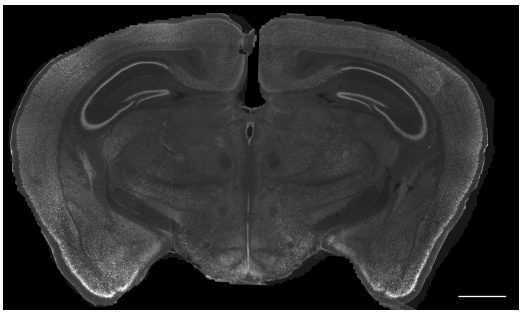
After finding the optimal rotation $R^*_{\alpha\beta}$, we apply the mapping method on full slices in A to find their corresponding full slices in the optimally rotated atlas. We then interpolate linearly on the matching slice indices to find the best matching slice for every other experimental slice in the experimental volume that is not selected in A.

## 3.3 2D to 2D Non-rigid Registration

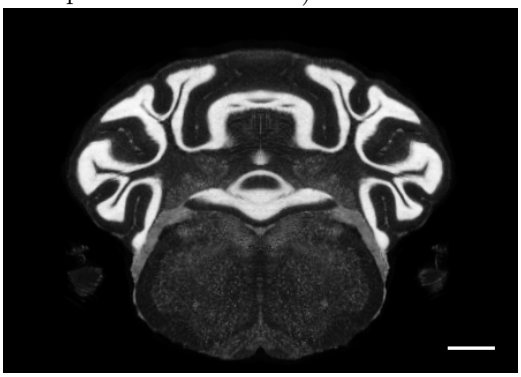
After the 2D to 3D registration, we map all the experimental slices to their computed corresponding slices in the optimally rotated atlas with the deformable registration and build a pixel-wise mapping from the 2D slice sequence to the atlas volume. We now need to register each experimental slice



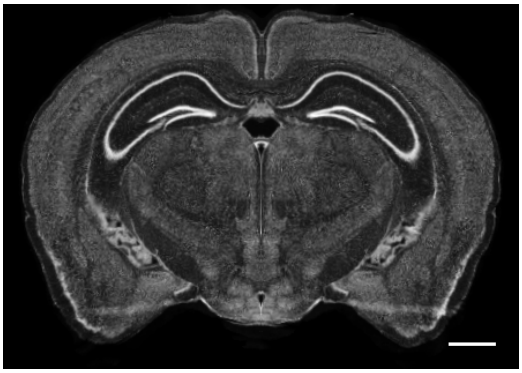
(A) Slice 15 in the experimental volume (ordered from posterior to anterior)



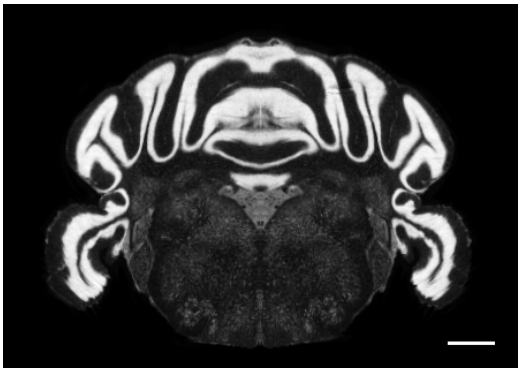
(D) Slice 85 in the experimental volume near the middle of the brain



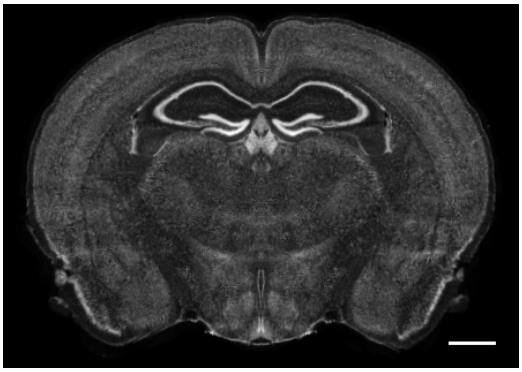
(B) Atlas (ordered from posterior to anterior) slice indexed at 55



(E) Atlas slice indexed at 224



(C) Atlas slice indexed at 85



(F) Atlas slice indexed at 257

Figure 3.1: Experimental slices and atlas slices with similar upper half and lower half to the experimental slices. Scale, 1 mm.

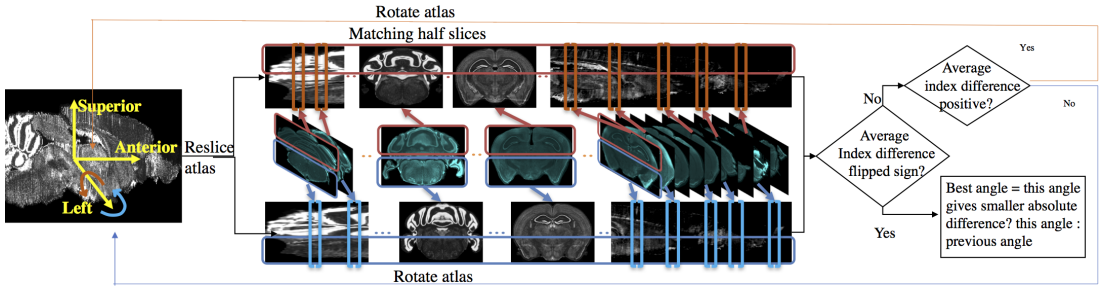



Figure 3.2: Flow chart for determining sectioning angle about the left-right (x) axis. In the matching half slices step, atlas is stretched for better illustration.

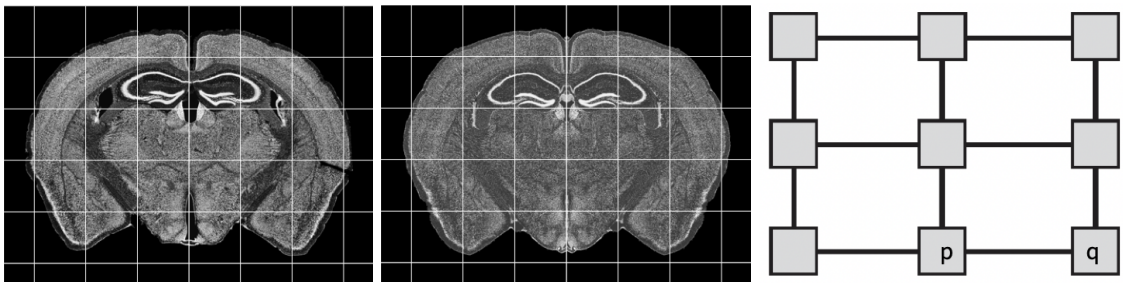
to its resliced atlas slice to build pixel-wise mapping. Since our experimental images are distorted, a deformable registration is needed. We first register the experimental slice to the reference slice with an affine registration as described in Section 3.1. After bringing the experimental slice to the coordinates of the reference image, we nonrigidly register the images to correct local distortions and potential structural difference. Let the globally transformed experimental image g be the target image, and its best matching resectioned atlas slice f be the source image, where $\Omega \subset \mathbb{Z}^2$ is the image domain. In the task of 2D registration, we aim at finding a transformation $\mathcal{T}$ such that

$$g(x) = f(\mathcal{T}(x)), \forall x \subset \Omega \tag{3.8}$$

where g and f become equivalent in terms of anatomical structures.

The most common approach to deformably register histological mouse brain images has been mutual information based free form deformation (FFD) [55]. Because of the properties of the experimental images described in Section 3.1, we continue to use the HOG difference as the similarity metric but with a smaller HOG cell size to fix local distortions. Since HOG is not differentiable, FFD is not directly applicable, and we need a method that does not require a derivative of the cost function. Therefore we build our work as discrete Markov Random Field (MRF) objective function [24].

Like in FFD, we superimpose a uniformly-spaced sparse grid $\mathcal{G} \subset \Omega$ on top of both images (Figure 3.3(a) and 3.3(b)). This uniform grid forms a 4-connected Markov random field model where a simple example is shown in Figure 3.3(c). For each node represented by a gray square in 3.3(c), we seek to assign a label that minimizes some overall cost function. In image registration applications, each label represents a specific horizonatal and vertical displacement of that control node. The cost function that is minimized is shown in Eq. 3.9 and consists of two terms: a unary term $V_p$ and a pairwise term $V_{pq}$. $V_p$ describes difference between of the HOG features between the patch in the experimental image centered at node $p$ and the patch in the reference atlas image centered at the same node. It represents how well the two regions match when the control point is moved by the




(a) Uniform grid on an experimental image.

(b) Same uniform grid on the reference atlas image.

(c) Example 4-connected grid model.

Figure 3.3: Uniform grid Markov random field model.

label's displacement. The pairwise term $V_{pq}$ is a smoothness term, and penalizes the label if its deplacement is different from the surrounding control node displacements. Each label $l_p$ corresponds to a displacement $d_p$ in a predefined displacement set $\mathcal{D}_p$. We define the bijective function $b_p$ between $L_p$ and $\mathcal{D}_p$ for each node $p$ as $b_p$: $d_p \rightarrow l_p$.

$$E_{MRF}(l) = \sum_{p \in \mathcal{G}} V_p(l_p) + \sum_{p \in \mathcal{G}} \sum_{q \in \mathcal{N}(p)} V_{pq}(l_p, l_q) \tag{3.9}$$

Direct implementation of the MRF-based image registration approach in [24] is simple. One only needs to use HOG difference to compute the unary term cost and keep the L1-norm regularization as described in the paper. However, since we are warping real brain tissues, we did more than that.

### 3.3.1 Model Elasticity with the Pairwise Term

The ventricular system spans throughout the brain, providing fluid pathways in the brain and creating regions of "empty" space in almost every histological slice as shown in Figure 3.4. Those cavities are easily deformed during slice preparation procedures and have much inter-subject variation. Thus when computing this MRF warp field, one needs to take into account the elasticity of different regions in the brain. By warping images to match with each other, we are essentially warping tissues: the more two adjacent control points are displaced, the more tension accumulates, if the two control points are connected through coherent tissue. In contrast, if they are separated by any hollow structure or empty space, the resulting tension is much smaller, which is why large deformations often occur to these regions.

In addition, the traditional and most common interpolation method for biomedical image analysis has been the B-spline model [55, 24], where each pixel is affected by $4 \times 4$ neighboring control points. In the case where two control points are separated by an empty space, a B-spline interpolation no longer makes sense because of discontinued tissue coherency. Therefore to better model tissue

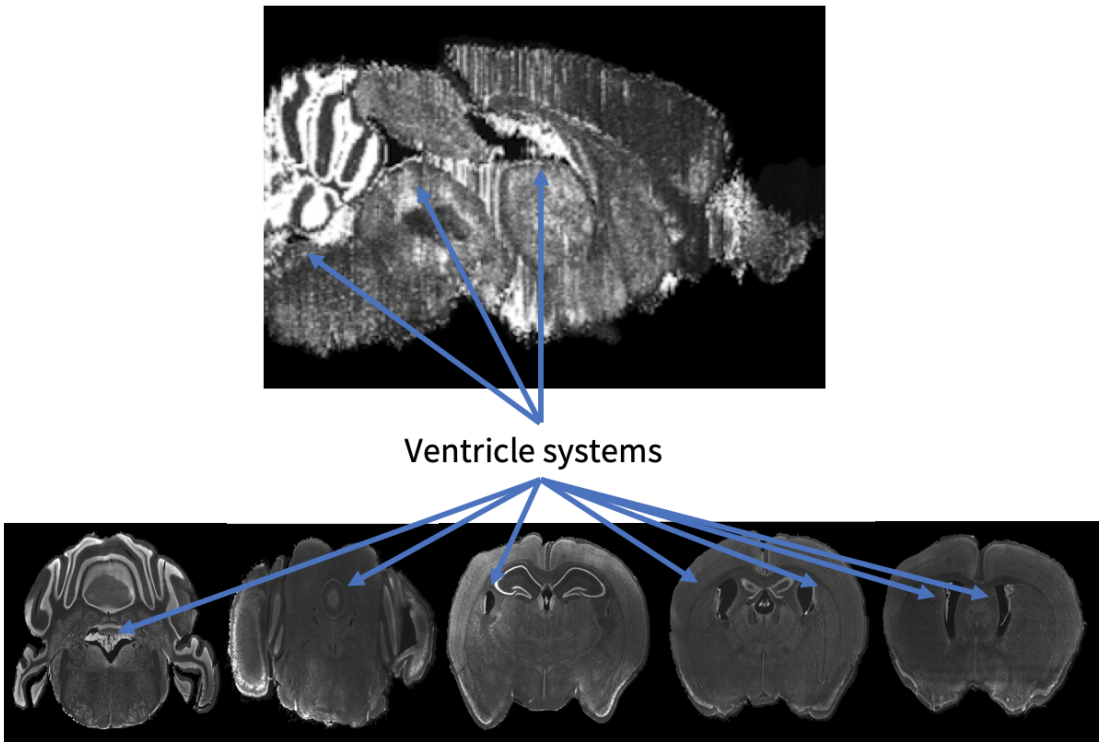



Figure 3.4: Ventricular system throughout the mouse brain. First row: sagittal view of the Nissl volume of ABA. Second row: example experimental slices in different sections of mouse brain.

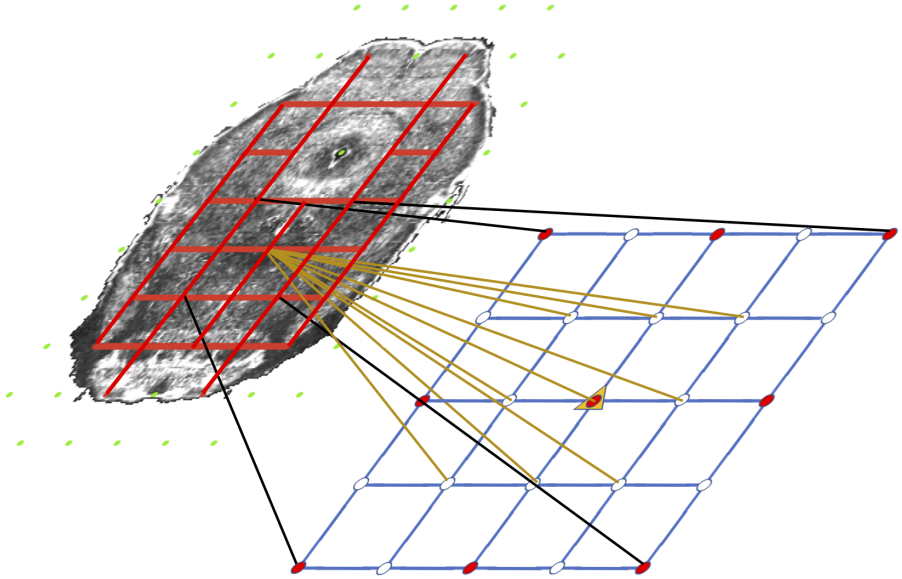


Figure 3.5: Illustration of the coherency model and grid refinement from level t to level t+1. Grids are overlaid on an atlas image (contrast adjusted for better illustration) to show the coherency model. Green nodes are free nodes which include nodes in the free regions — ventricle systems and background — and affect real tissue. Red nodes are coherent nodes which cannot be seen since they become part of the tension edges, represented by red line segments between coherent nodes. During refinement, image resolution is 2× the resolution of last level. The grid spacing remains the same. Therefore the grid quadruples in each direction, which is shown in the lower grid. The motion of existing nodes are carried onto the next level. The motion of non-existing nodes in the lower grids are interpolated from the motion of the existing nodes.

deformation, we use the simple bilinear model where a point is only affected by its direct $2 \times 2$ neighboring control points. Of course, now our warp generation needs to ensure some smoothness.

Our system does so with a very simplistic model. We divide each target slice into two regions: free (ventricular system and background), and coherent (other areas) based on the annotation volume of ABA. We then classify the nodes as coherent (red) or free (green in Figure 3.5) based on if they are inside a coherent or a free region. The idea is tension only accumulates when we compress or stretch two nodes that are connected solely with a coherent region. If there is an empty space between two nodes, intuitively compressing them or stretching them should not build tension in between. Based on this property, we use the pairwise term — the traditional regularization term — to model tension accumulated between nodes with which we stretch or compress brain tissues. There is an edge (red line segment in Figure 3.5) only if the connecting line segment between the two nodes only crosses coherent regions which indicates both nodes must be coherent as well.

We first extract mask images $r_c$ and $r_e$ representing coherent region — tissue — and empty space

region — ventricles and background — respectively from the reference annotation and project them to the source image. We further group control points as coherent or free. Coherent control points are inside coherent regions. Free control points are the control points inside an empty space, and moving the control point will affect pixels inside coherent regions. Formly the classification of the control nodes are given in Eq. 3.10:

$$p \in \begin{cases} \mathcal{G}_{coherent}, & \text{if } r_c(p) = 1 \\ \mathcal{G}_{free} & \text{if } \sum_{x \in \Omega}[\eta^{-1}(|x-p|) \cdot r_c] > 0 \\ & \text{and } r_e(p) = 1 \end{cases} \tag{3.10}$$

where the inverse influence function, $\eta(.)^{-1}$, adopted in [24], masks pixels influenced by a control point p. We include the influence function in control points classification, because we only care about control points that affect image appearance.

We further define a tension edge set, $\mathbf{E}$, where tension accumulates when moving the two control points connected by an edge in this set. Basically an edge $e_{pq}$ is in $\mathbf{E}$ if the line segment connecting Node $p$ and Node $q$ only crosses coherent region $r_c$. Because the spring potential energy is proportional to the square of displacement, we use squared difference as the pairwise term:

$$V_{pq}(l_p, l_q) = \lambda(d_{l_p} - d_{l_q})^2, e_{pq} \in \mathbf{E} \tag{3.11}$$

where $\lambda$ is a regulation parameter.

### 3.3.2 Multi-level Estimation

We need to be able to both correct large distortions and make small changes to achieve good results. For both computation efficiency and quality of results, we use a multilevel approach. Since we are trying to model the tension that the deformations create, we need the pairwise energy terms to accumulate as we refine the grids. This requirement means that we cannot use the approach used by Glocker *et al.* [24], but instead solve the problem using a method where each refinement level maintains knowledge of the distortions created by previous levels.

The conventional multilevel approach [24] repeats the same procedure with progressively finer grids: locations of the control grid points that minimize the unary and pairwise terms are computed, and then the resulting image is warped to match these new grid locations. The next level grid is added to the warped image, and the process is repeated. To maintain tension in a realistic way, we do not reset the grids and tension after each iteration and use each iteration to simply update the allowable possible positions (labels) for the next iteration. More formally, to carry the squared form tension correctly to the next level, at every level $t$ for each node $p$ we update the possible discrete displacements $\mathcal{D}_t^p$ to reflect the accumulated prior displacements of the node plus the current

displacement to be evaluated at the current level.

To correctly carry on calculated displacements to the next level, we first need to compute the set of possible locations for each grid point, which depends on the results from the prior level. To do this, we denote the grid at level t as $\mathcal{G}^t$ and the influence function as $\eta^t$. Let $d^t_{l_p}$ be displacement at node $p$ with label $l$ at level t. At each level t, we estimate best motion $d^t_{l_p}$ at each node $p \in \mathcal{G}^t$ and bilinearly interpolate them to get the initial displacement for each node in $\mathcal{G}^{t+1}$ at the next level. We denote this preset displacement at node $p$ as $\bar{d}^{t+1}_{l_p}$, therefore the set of possible displacement at node p level t+1 is given by the sum of this preset displacement and possible displacement in level t+1:

$$\mathcal{D}^{t+1}_p = \{d : d = \bar{d}^{l_p}_{t+1} + \theta, \theta \in \Theta\} \tag{3.12}$$

where $\Theta$ is the allowable additional displacements and is the same for every node.

Having created a set of possible locations for each grid point, we next need to create the image that we will compare at this level to compute the similarity. In previous work, this warped image is input to this level, but we need to compute the image from the displacements of the previous level's control points and the labels associated with the node we are evaluating. When estimating the local patch around node $p$ at level $t+1$, we use the positions of the grid points that are around p from the prior level and the position of $p$ for the given label at this level. This provides an estimate that incorporates the warp from the prior level and an estimate of the additional warp created by moving p to the position indicated by this label at the current level. For simplicity and computational efficiency, this estimate ignores the warp that will be caused when other control points at this level move.

We denote the patch that is affected by p in the first level function $\eta_0^{-1}(|x-p|)$ as $\mathcal{R}_{0,p}$. The control points in the patch at level t+1 is defined as:

$$\mathcal{N}^{t+1}_p = q \in \mathcal{G}^{t+1} : \eta_0(|q-p|) > 0 \tag{3.13}$$

To create the image that we will compare, we set the nodes in $\mathcal{N}_p$ at the values from level $t$, except for node $p$ which we evaluate with the displacements from the current level in the set $\mathcal{D}^{t+1}_p$. Therefore the transformation applied to the affected region $\mathcal{R}_{0,p}$ when we associate label $l_p$ with node $p$ at level $t+1$ is:

$$\begin{aligned}\mathcal{T}^{t+1}_p(x, l^p) = x + \sum_{q \in \mathcal{N}^{t+1}_p, q \neq p} & \eta(|x-q|)d_{l_q} \\ & + \eta(|x-p|)d_{l_p}\end{aligned} \tag{3.14}$$

Thus the unary term is given by the similarity measure between the warped patch and target patch:

$$V^{t+1}_p(l_p) = \rho(g(x), f(\mathcal{T}^{t+1}_p(x, l_p)), l^p \in \mathcal{D}^{t+1}_p \tag{3.15}$$

where $x \in \mathcal{R}_{0,p}$ and $\rho$ measures the difference score between two images. Since at every level, the only region that changes when we are approximating the change for each possible label is centered around the node being evaluated, we approximate this change by simply translating the patch centered at the approximate new node coordinates $p + d^{l_p^t}$ with possible additive displacements $d^{l_p^{t+1}}$ at the current level and compare the HOG similarity with the patch that is centered at the node's original coordinates in the fixed image. Therefore the unary term is estimated as:

$$\overline{V}_p^{t+1}(l_p) = \rho(g(\mathcal{R}_{t,p}), f(\mathcal{R}_{t,p+d^{l_p^t}+d^{l_p^{t+1}}})) \tag{3.16}$$

where $\mathcal{R}_{t,p}$ denotes the patch centered at node $p$ in the fixed image, and $\mathcal{R}_{t,p+d^{l_p^t}+d^{l_p^{t+1}}}$ denotes the patch centered at newly estimated coordinates of displaced node $p$.

Eventually we formulate the MRF energy function at level t as the normalized unary similarity term to the corresponding atlas image and the pairwise term:

$$E^t = \sum_{p \in \mathcal{G}^t} \mathcal{F} \circ V^t_{p,atlas}(l_p) + \sum_{p \in \mathcal{G}^t} \sum_{q \in \{q : e_{pq} \in \mathbf{E}^t\}} \mathcal{F} \circ V^t_{pq}(l_p, l_q) \tag{3.17}$$

where $\mathcal{F}$ denotes the normalization operation. We obtain the HOG difference for all nodes and all labels at each level and normalize the matrix so that the values are within range $[0, 1]$. The pairwise term is normalized in the same fashion.

Because sometimes some features in the atlas do not show up in the experimental volume, while the features in the same experimental volume are consistent, we add an additional unary term to the energy function in order to encourage consistency of warped experimental slices. This term encourages features in the target experimental image to align with the warped experimental image before it. Therefore energy function in Equation 3.17 now becomes:

$$E^t = \sum_{p \in \mathcal{G}^t} \mathcal{F} \circ V^t_{p,atlas}(l_p) + \sum_{p \in \mathcal{G}^t} \mathcal{F} \circ V^t_{p,prev}(l_p) + \sum_{p \in \mathcal{G}^t} \sum_{q \in \{q : e_{pq} \in \mathbf{E}^t\}} \mathcal{F} \circ V^t_{pq}(l_p, l_q) \tag{3.18}$$

where $V_{p,prev}$ represents the unary term between the current experimental image and the warped experimental image before it. This term is zero when the current experimental image is the first slice in Set A.

Because the free nodes are not constrained with any pairwise terms, they are essentially assigned labels that minimize the unary terms:

$$l_p^t = \underset{l \in b(\mathcal{D}_p^t)}{\arg\min} V^t_{p,atlas}(l_p) \tag{3.19}$$

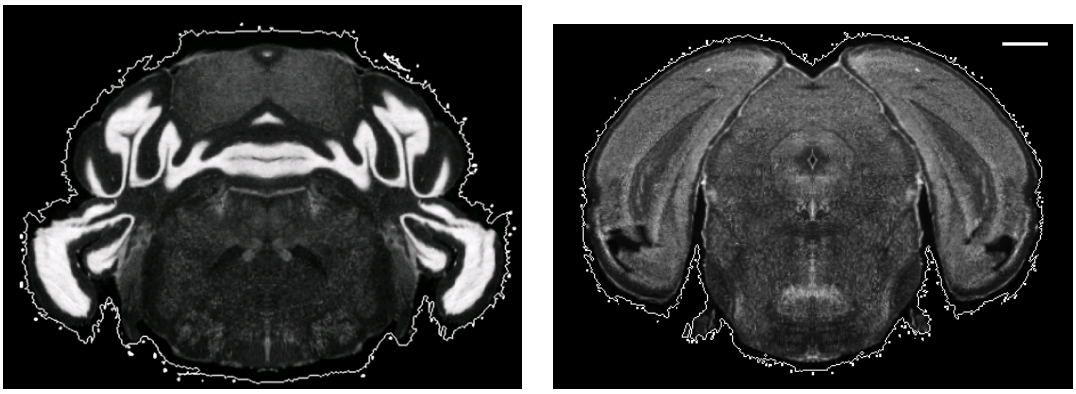


Figure 3.6: Example atlas slices with nonzero intensity regions circled by white contours. Scale, 1 mm.

where $p \in \mathcal{G}^t_{free}$. Coherent labels are solved by minimizing the energy function:

$$E^t_{coherent} = \sum_{p \in \mathcal{G}^t_{coherent}} \mathcal{F} \circ V^t_{p,atlas}(l_p) + \sum_{p \in \mathcal{G}^t_{coherent}} \sum_{q \in \{q : e_{pq} \in \mathbf{E}^t\}} \mathcal{F} \circ V^t_{pq}(l_p, l_q) \tag{3.20}$$

### 3.3.3 Contour Alignment with Symmetric Difference

While HOG matches internal features well, we find it hard to align the contour of images. The atlas has very low-intensity pixels around real brain tissues as shown in Figure 3.6. Because HOG's relative insensitivity to intensity, this noise can cause errors in contour alignment. We added an intensity threshold in the process of computing HOG — if a pixel's intensity is lower than the threshold, its gradient is not included in the histogram. However, this does not solve the problem: a single threshold intensity cannot eliminate the background noise perfectly because of uneven illumination and the presence of regions having few stained neurons. To remove the unwanted background, we make use of the fact that the corresponding slice's annotation in the annotation volume of brain atlas is conservative: it is inside the actual imaged tissue. Thus all pixels inside the annotation are foreground pixels. We further segment the pixels that are annotated as background in the annotation to either the foreground or the real background using the intensity information of the image. We build an energy function so that the unary terms try to minimize intensity difference among the pixels that are in the same class, and the pairwise term encourages two neighboring pixels to be grouped in different classes when the intensity difference is large and to be grouped in the same class when the intensity difference is small:

$$E(f) = \sum_{p \in \mathcal{V}} \mathbb{1}(f_p = 0) \cdot i_p + \sum_{p \in \mathcal{V}} \mathbb{1}(f_p = 1) \cdot \mid i_p - I_{average} \mid + \sum_{p,q \in \mathcal{E}} e^{-(i_p - i_q)^2} \tag{3.21}$$

where $\mathcal{V}$ represents all the pixels to be classified - the pixels are annotated as background in the annotation, $\mathcal{E}$ denotes pixels that are in the 4-connected neighborhood, f denotes the assignment of background — 0 or foreground — 1, $i_p$ represents the intensity at pixel $p$, and $I_{average}$ is the average intensity of all nonzero pixels in the image. Solving this energy function, we can obtain satisfactory result except that some very dark tissue regions near a slice's contour will be removed in some slices. To fix it, we keep the otherwise removed regions if the area is well-connected with its surrounding regions. This is accomplished by morphological eroding and dilating the to-be-removed regions with a disk of 20 pixels. We keep a region if it survives the opening operation. Same numbers are used across slices. We fill in holes in the computed mask so that the mask consists of a single piece.

After fixing the background noise, we find it still hard to use HOG difference to align image contours because HOG difference reduces dramatically only when after transformation the contours overlap or are separated by a distance smaller than the HOG cell size. Moreover, since the atlas is not a smooth volume, after rotation, the contours may be jagged — creating unwanted gradients. A more sensitive and more robust metric is needed.

When displacing nodes that affect image contours, we are essentially warping the contour to maximize the overlapped region of the two images or equivalently to minimize the symmetric difference of image foregrounds. Therefore, if a node $p$ influences contour pixels, we modify the HOG difference term to its corresponding atlas patch to be the symmetric difference of the two tissue masks. The goal is to make the contour of the two tissues align. Because node $p$ is essentially a free node, Eq.3.19 is updated to:

$$l_p^t = \underset{l \in b(\mathcal{D}_p^t)}{\arg\min} \sum_{\eta^{-1}(|x-p|)=1} (m_f(x) - m_g(x)) + V_{p,prev}^t(l_p) \text{ if } \sum_{x \in \Omega} \eta^{-1}(|x-p|) \cdot c_e > 0 \tag{3.22}$$

where the contour of the experiment image's real tissue is denoted by $c_e$, and the real tissue in image $f$ and $g$ are $m_f$ and $m_g$. In Eq. 3.19 we estimate the warped patch by translating the patch center. This estimation improves computation speed while retaining performance when the similarity measure is HOG difference or another metric that involves more internal information. The shape of the experimental images is often deformed in the preparation process. However, simple translation does not change the shape. It only reduces the disjoint area but cannot find a transformation that reverts the deformation. Therefore instead of simply translating the patch, we warp the binary masks to evaluate this symmetric difference term.

## 3.4 Pixel-wise Mapping Improved

Initially this non-rigid registration approach looked like it was working well. The resulting slice and atlas data matched well. However when we mapped images with a stitching artifact in them, we noticed a problem with our current approach. The stitching artifact came from the scanning

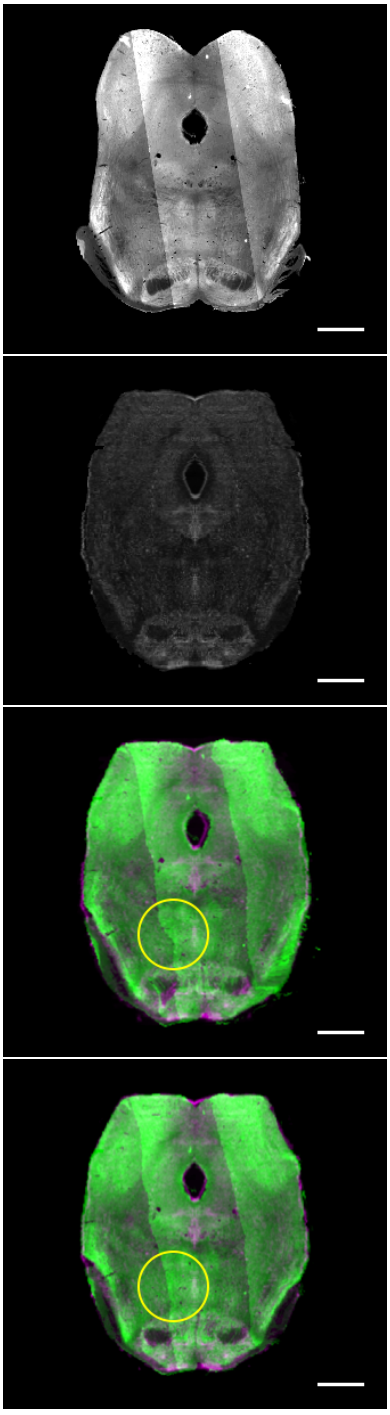
microscope used to generate the images. For these images the scope had a problem and did not scan with the same intensity when it created the experimental slice image. This difference generated two scan lines in the resulting image, as shown in Figure 3.7(a). What we see is that in a slice we thought should be smoothly distorted has bumpy curves shown in Figure 3.7(c) which seemed artificial. Our guess is that the problem is some of these control nodes are in relatively uniform regions. Because histological images are noisy, the noise signal may be stronger than the regularization signal that moves the nodes according to its neighborhood motion. So to experiment and to confirm this was the problem we manually removed the control points in places that did not have salient features. When we did that, the results are much better, as shown in Figure 3.7(d). The location the control points was the cause of the problem.

In order to solve this problem automatically. We need to develop a more sophisticated control point generation algorithm. The first question is where these control points should be. The goal of aligning two biomedical images is to place the corresponding biological features at the same coordinate. After aligning those corresponding features, the relatively uniform-intensity regions' new coordinates should be interpolated based on the coordinates of the more salient features. From this experiment, we learned that aligning uniform regions is the cause of the bumpy curve problem. Of course, there are relatively uniform-intensity regions with significant biological features, and these pixels are hard to align. Therefore, the new control points should be placed on the boundary of the significant biological features where there are visual contrast to surrounding regions. In this section, we describe our approach to provide an improved pixel-wise mapping method based on generating salient control points.

### 3.4.1 Salient Point Extraction

Our first task is to extract the contours of contrasted regions or biological features. Because the ABA contains an annotation volume, we planned to use the contours of biological features in the annotation volume directly. However after overlaying the same plane in the annotation volume to a reference Nissl image, we find the annotation volume is inaccurate as shown in Figure 3.8. Displayed in the figure is an unrotated atlas slice taken directly from the atlas and its corresponding annotation label boundaries. One can easily see many inconsistent labeling especially at the contour of biological features. We circled two visually significant ones. In rotated atlas, the labeling would appear worse because the labels were completed manually, and no smoothness was enforced across slices. Therefore we need to extract real contours.

We tried different edge extraction methods including Sobel, Prewitt, Roberts, Laplacian of Gaussian, and Canny edge filters. Sobel, Prewitt, and Roberts filters generate very similar results with main edges extracted but with a lot of noise points. Laplacian of Gaussian and Canny edge filters are very sensitive to noise and interpret some noise signals as real edges. Extracted edges consist of many fake edges where the edges are strong noisy points, or contours of small noisy blobs.



(a) An experimental slice with stitching artifact. Contrast adjusted because original image is too dim. Affinely registered to the corresponding atlas slice.

(b) Corresponding atlas slice of (a).

(c) Experimental slice nonrigidly registered to the corresponding atlas slice. Figure shows an overlay of the nonrigidly-warped experimental image (green) and the reference atlas image (magenta). Bumpy curve circled in yellow.

(d) Overlay after removing control points in relatively uniform regions. Nonrigidly-warped experimental image (green), reference atlas image (magenta). Original bumpy curve location circled in yellow.

Figure 3.7: Comparison of results with uniform grids and control points often around significant features of a selected experimental image. Scale, 1 mm.

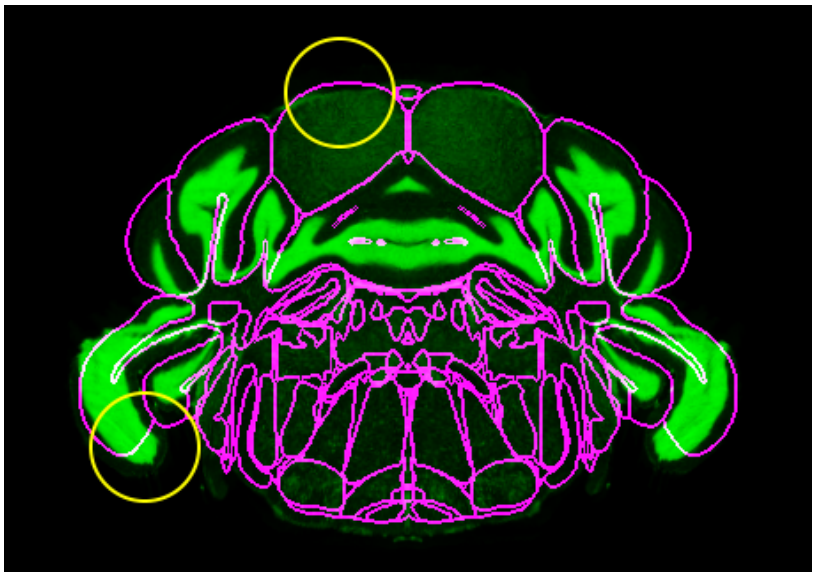


Figure 3.8: Overlay of an unrotated atlas image (green) and its corresponding annotation contours from the annotation volume (magenta). Two examples among all inconsistent labels are circled in yellow. Scale, 1 mm

It is hard to eliminate those fake edges by tuning the parameters when using these two methods. We show results with Sobel filter and the Canny edge method in Figure 3.9. The former more traditional filters, for example Sobel, generate closer to ground truth edges than the more advanced filters.

Because we have annotations that are close to ground truth, we make use of this information to remove the fake edges in the extracted edges based on the observation that real edges are mostly closer to the annotation boundary than fake edges. Because real edges are very likely to be stronger than fake edges, we also make use of the gradient information. We did it by scoring all potential edge points with a scoring function. The score function $v$ consists of two terms: a distance term that is based on a point's distance to the closest annotation boundary and a magnitude term that is based on the gradient magnitude at the point.

We model the distance term as a Gaussian function and denote it as $v_1$. The term is greatest when it is exactly on top of an annotation boundary. It attenuates as the distance from the point to the closest annotation boundary gets larger. We first extract the annotation boundaries, a binary image, and then compute the distance transform, [47], of this binary image that measures a pixel's Euclidean distance to its closest non-zero pixel. We denote this distance as $d$. Because 99.7% of values drawn from a normal distribution are within three standard deviations, and we want the points that are farthest to annotation boundaries to be close to zero, we make the standard deviation $\sigma$ to be one third of the longest distance from all potential points to their closest annotation boundary. The distance term is given by $v_1 = e^{-d^2/(2\sigma^2)}$.

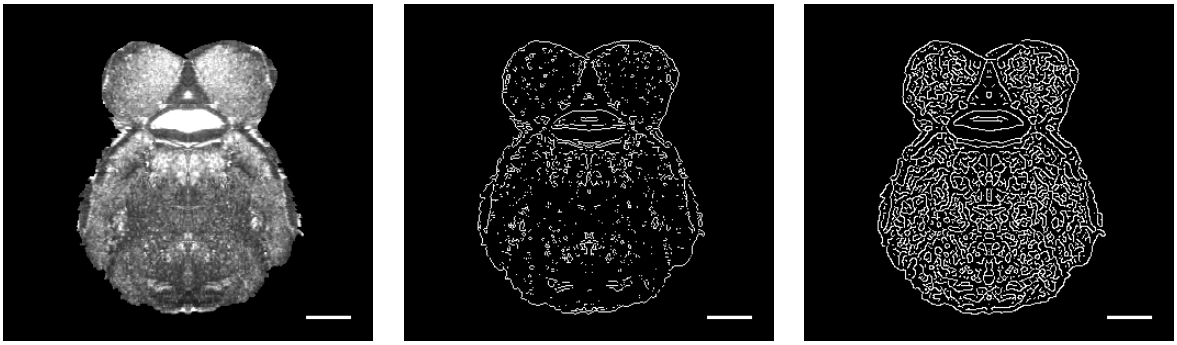


Figure 3.9: Left: original experimental image (contrast adjustment for better visualization). Middle: Edges found with Sobel approximation to the derivation. Right: Edges found by looking for local maxima of the gradient with the Canny method. Scale, 1 mm.

Therefore the score function is straightforward. We normalize the magnitude $g$ so that the range of the gradient term is $[0, 1]$, the same as the distance term $v_1$. The score function is given by $v = g * v_1$.

### 3.4.2 Sparse Selection

Now we are able to rank the potential edge points with the score function $v$. We cannot directly choose the first 10, 20, or 50 points, because we want the control points to spread out, cover most of the regions while keeping a minimum distance between each pair of control points to avoid repetitive computation. We originally set the minimum spacing to be the spacing in our uniform grid approach in Section 3.3. However, because now all control points are inside real tissues, we have less effective areas, setting the same spacing sometimes ends with very few control points in the coarsest iteration level. As a result, we set the minimum spacing to be half of the original spacing.

We first rank the potential points from the highest score. We directly keep the most probable point, compute the Euclidean distance from other potential points to this most probable point, and delete all the points that are closer than the minimum spacing from the selected points. We keep this process until no points can be selected. For most of the images, this scheme brings us satisfactory control points that are only on the contours of contrasted regions (including the contour of the image) and have a good distribution across the image. However there are still some images where control points are generated in relatively uniform intensity regions where no significant features are around.

### 3.4.3 Cycle Consistency

The origin of these bad control points are from fake edges near some annotation contours, generally near annotation boundaries in regions of low contrast. For example, the superior colliculus, a layered,

multi-sensory structure, is further divided into many subregions that are in charge of processing multiple signals, while there is no significant intensity difference between these regions. Since different brain regions have different level of subregions which may or may not be divided based on intensity difference, to avoid this issue requires manual reencoding to exclude the regions that present no visual contrast to their neighboring regions.

We ended up taking a different approach, which required much less manual effort. We observe that good control points have image patches that match corresponding image patches in the experimental slice (that is what makes them a good control point), while the bad control points do not. Thus we can see if we are consistent when we map from the control point to a point in the experimental slice, and then map from the point in the experimental slice back to the atlas. If we end up at the same control point, we keep the point and say it is cycle consistent. This idea is similar to the four-step procedure check recommended in [11]: original → target → target check → original, as well as the "cycle consistency" given in [70] for unpaired image-to-image translation.

We add this step in the sparse selection described in Section 3.4.2. While there are still candidate points left, we choose the point with the highest score given in Section 3.4.1. Instead of choosing this highest-score point directly, we go ahead and evaluate its cycle consistency. We take the patch with height and width given by spacing constraint, center this patch at this candidate control point and compute this patch's HOG description. We denote the coordinate of this candidate control point as $x_n, y_n$. Remember we do not warp the moving image until after the last iteration, during each iteration, we take into account the previously computed best motion of this control node $x_m, y_m$, then the corresponding coordinate of this control node in the moving image would be $x_n + x_m, y_n + y_m$. We then adjust the center of this patch centered at the corresponding coordinate in the moving image, and compute the best correspondence this this iteration level. Therefore, with additional displacement, we check every same-size patch centered at $x_n + x_m + x_a, y_n + y_m + y_a$ for different additional motion $x_a$ and $y_a$ at the current level. The additional motion are bounded by half of the spacing we choose for this iteration. We describe the patches in moving image with HOG descriptor as well. With a similar idea in [24], we only directly look at patches centered at different coordinates, but do not actually warp them. The best matching patch is the one that gives the minimum HOG difference from the original patch in the fixed image, we denote the coordinates where this best matching patch is centered as $x_n + x_m + {x_a}^*, y_n + x_m + {y_a}^*$.

From this best corresponding point in the moving image, we repeat the process of finding the best matching point back in the fixed image with a similar process. Now we view the patch centered at the best matching point in the moving image as the fixed patch. In previous iterations, we have computed the displacement of this control node and describe it with the HOG descriptor. Because the control point in the fixed image has a previous computed displacements of $x_m, y_m$, assuming this best matching point in the moving image can successfully map back to the original coordinate in the fixed image, we add the relative displacements which is now $-x_m, -y_m$ to $x_n + x_m + {x_a}^*, y_n + x_m + {y_a}^*$

and start searching best matching patch in the fixed image from there. The searching space of additional displacements is the same as before. Again we compute the additional displacements that give the minimum HOG difference. If the additional displacements is exactly $-x_a{}^*, -y_a{}^*$. We say the candidate control point is cycle-consistent and keep the point. We repeat the sparse selection 3.4.2 with additional cycle consistency check for every highest-scoring candidate point until no point is left. We do not check cycle consistency on the boundary points.

### 3.4.4 Pairwise Term with Salient Control Nodes

We previously described our pairwise term that models elasticity in Section 3.3.1 where we adopt the uniform grid from the freeform deformation [55]. With salient control nodes, we cannot directly use the previous approach, because:

1. The salient control nodes no longer form a uniform grid, finding neighbors is no longer trivial.
2. All salient control nodes are now inside a tissue or on the boundary of a tissue. Therefore there are only coherent nodes.
3. Previously all neighboring control nodes were separated with a same distance — the spacing. The edge energy could be generated using the same "sping" for all edges. Now nodes are separated with different distances, and the tension term need to be modified accordingly.

We still want a regularization term that smoothes the displacements of two adjacent nodes. We can achieve this by partition the image plane based on the salient control points. Specifically we use the most commonly used meshing method — Delaunay Triangulation [37, 15]. As before, we still do not want to penalize difference in motion between two adjacent nodes separated by an empty space. Therefore after building the triangulation, we simply check every generated connection and remove the edge if two nodes are separated by an empty space as given by the annotation image. This empty space can also be obtained with a threshold in most images. For harder cases, a segmentation network needs to be learned in order to detect empty spaces automatically. Previously we viewed the edges as using the same spring, now because the length of edges are variable, we view them as cascaded unit springs where the number of springs is different for each edge. When an edge is compressed or stretched, the deviation from its original length is shared by each unit spring. Suppose the number of unit springs that form a specific edge is given by $n$, and the deformation length is given by $d$, the deformation of each unit spring would be $d/n$. Therefore the spring energy of each unit spring is $((d_p - d_q)/n)^2$, and the tension built on the edge is $(d_p - d_q)^2/n$, where $d_p$ and $d_q$ are displacements of two nodes connected through an edge. We modify Equation 3.11 accordingly.

## 3.5 Mapping Neurons with the Deformation Field

The end goal of the this mapping is usually locating neurons of interest in the atlas, and from the information gained from this mapping learn about neural connections in the brain. This task requires one to first identify these neurons in the experimental volume, which can be done manually or automatically. After obtaining these locations of neurons in each experimental slice, with the transformation fields obtained from our program — atlas rotation, best matching corresponding slices, affine transformation field, and nonrigid deformation field, we compute the location of these neurons in the atlas and can aggregate data from many experimental brains.

Given a neuron coordinate $C$ in an experimental slice, the original location in the atlas $C'$ is given by:

$$C' = \mathcal{S}^{-1}\mathcal{D}\mathcal{T}C \tag{3.23}$$

where $\mathcal{T}$ is the affine transformation matrix that aligns the experimental image globally to the corresponding resectioned atlas, $D$ is the deformation field obtained with our nonrigid registration. These two transformation map the experimental points to the coordinate system of the corresponding resliced atlas image. $\mathcal{S}$ includes the rotation and the slicing that defines the corresponding plane, and $\mathcal{S}^{-1}$ maps the points in the corresponding plane back to the atlas coordinate system.

## 3.6 Implementation

### HOG Cell Size

We use a cell size of 15 pixels to measure the image similarity. This relatively large cell size allows us to capture structural similarities even with uncorrected small distortions. For nonrigid registration in Section 3.3, we decrease the cell size to 4 pixels, because the purpose of this step is to correct distortions. In both steps, the block size is $2 \times 2$. HOG is computed with the Vlfeat toolbox [61].

### Set A

We select a subset, A, from all the slices I to find the best cutting angle and the best corresponding slices. Creating a subset provides more freedom for the slice matching algorithm and provides better results. It also allows us to remove bad slices from this part of the process to improve the quality of results. The extra freedom can be understood by looking at Equation 3.3 which gives the spatial constraint. The neighborhood in terms of number of slices we would search is given by $2 * \theta * S_E/S_A(A_i - A_{i-1})$. Because $\theta$ is usually selected to be 20%, the thickness of experimental slices $S_E$ is usually twice the resolution of the atlas $S_A$, if we select adjacent slices $A_{i-1}$ and $A_i$, the neighborhood of slices we would search for slice $k$ would only be 0.8 slices. A slice index can only be integer, even if we round the slice numbers, we would end up only restraining the range to one

slice. There would be no optimization at all. For our simulated full brain data, which contains about 250 slices, we use every sixth slice. For our sectional simulated data, which contains about 50 slices, we used every third slice. For good-quality experimental brains, we recommend similar automatic selection of these slices as well. For brains with relatively more damaged slices, we recommend checking the automatically selected slices and replacing the index of very damaged slices with a nearby good-quality slice. A good selection of subset A is a sparse selection that covers the full or near-full range of all slices while avoiding most damaged slices.

### Parameter Selection in the Energy Function

All terms — the unary terms and the pairwise term — in the energy functions presented in this paper are normalized to the range $[0, 1]$. Since the experimental images and the atlas volume are of the same modality, and the terms are normalized, the parameters before each term do not need to be heavily tuned to yield good results. Our original energy function consists of only one unary term — the HOG similarity term to the atlas slice — and the pairwise term. With several experiments, we find equal weight between the unary and the pairwise term generates the best visual result. In the general energy function in Eq. 3.17, the HOG similarity term to the previous slice is added to encourage smoothness in the "reconstructed" volume and make the features that do not exist in the atlas but exist in the experimental volumes more consistent. We add an additional Euclidean distance term between the two aqueduct contour point sets in Eq. 3.18 to suit our dataset better. Since the HOG difference to the corresponding atlas slice is the dominant term, we set it to be three times as strong to the HOG difference term to the previous slice and the Euclidean distance term between the two aqueduct contour point sets in both forms of the energy function. The coefficients before the pairwise term is set to be the sum of the previous coefficients to maintain the equal weight between unary and pairwise terms.

### Iteration Details

In order to correct distortions of various size, we employ a hierarchical sets of displacement vectors by downsizing the image pairs to different extent. We use three iterations to complete the 2D nonrigid registration described in Section 3.3 and in Section 3.4. In the first iteration, we downsample both images $4\times$ in both horizontal and vertical directions. In the second iteration, images are downsampled by $2\times$. In the final iteration, we use the original resolution. The grid spacing is $16 \times 16$ in all iterations in Section 3.3. We use the same spacing of 16 in Section 3.4. In Section 3.4, when selecting candidate points, if a candidate point passes the cycle consistency test, we delete all other candidate points within distance of one half of the spacing. We generate a search space for each optimization sweep by defining a maximum range of displacements along the horizontal and vertical directions. The maximum displacement at each level is set to be half of the grid spacing in both Section 3.3 and 3.4. Therefore the total number of labels are $17 \times 17$ in each iteration. After an optimization sweep,

we rescale the displacements to the scale of the next iteration and take this information into account when optimizing the next iteration. This optimization sweep is repeated for the number of iterations, after which we apply the accumulated displacements to the moving image. The optimization is computed with tree-reweighted message passing (TRW), more specifically TRW-S [13, 35].

## Caching

In the improved 2D nonrigid registration in Section 3.4, we find the computation time is around twice the original 5 minutes per pair of slices, even when we have paralleled the computation searching for the best matching patches centered at surrounding points. This was initially surprising, since in the improved version, we have fewer but better control points. The increase in runtime was caused by the cycle consistency check. If a candidate control point is not cycle-consistent, we simply toss it without deleting any neighbor candidate points. We keep on checking for cycle consistency from the highest-scoring point until one satisfies cycle consistency when we can delete all the points within a neighborhood given by one half of the spacing.

Since most candidate points are connected and form an edge or are near one other, when searching for a best matching patch with different additional displacements, we repeatedly compute a HOG score for the same patch many times, repeating the computation each time a close neighbor point is tried after the prior point fails its cycle consistency test. In order to save computation time, we save the computed HOG description vectors of patches in the moving image into a cache. We build a cache matrix and save the HOG descriptor vector as rows indexed by the corresponding pixel's row and column. We save the HOG description of an image patch centered at a pixel to its corresponding place in the matrix. Because in the cycle consistency check, we do not warp an image, but slide a window around the point to estimate the difference, the HOG description of an image patch centered at a specific coordinate is unique and can be reused during the same level where the images are sampled at the same rate. If a patch's HOG description is previously computed, we read the description directly from the cache. Otherwise we compute it and save to the cache.

Because we need to write the HOG descriptor to the cache when computing for different displacements, we cannot write to a shared matrix/cache if this computation is parallel (this is an internal MATLAB restriction). Therefore we can only consider these two options separately rather than combining them. The computation time with caching is 3 to 4 minutes and with parallel computation it is around 10 minutes per image pair. This caching approach our new approach is better than our uniform grid method which takes around 5 minutes per image pair.

We measured the time to compute the corresponding point of the first two candidate points that are evaluated to understand the time performance better. The results are summarized in Table 3.1. We can see for this specific image pair, in all iterations, the computation time to find the best matching point of the first point is slightly larger than the computation time of the same first point without parallel computation. However, this time reduced to about a half when computing the best

| Method | Cache | | Parallel | |
|---|---|---|---|---|
| Iteration | First Point | Second Point | First Point | Second Point |
| 1 | 0.19 | 0.07 | 0.14 | 0.14 |
| 2 | 0.18 | 0.08 | 0.12 | 0.13 |
| 3 | 0.15 | 0.09 | 0.14 | 0.13 |

Table 3.1: Computation time, Cache vs. Parallel.

matching point of a second candidate point, which brings the computation time below the time it would take with parallel computation. In all three iterations, from the cache method time, we can conclude the second point is near the first point and reused HOG descriptions in the cache. From this example, we can see the caching method is faster because re-usage of HOG description happens often.

## 3.7 Evaluation on Simulated Datasets

Because of the lack of ground truth in real experimental brains, and experimental brains may be sectional brains or full brains depending on the study, we generated both a sectional brain and a full simulated brain from the atlas as our simulated dataset with known transformations. The simulated sectional brain is generated from the hindbrain section in the atlas. It also reflects the corresponding section in the study [53]. Both the sectional brain and the original full atlas are resliced with sectioning angles different from that of the original atlas. The most anterior and posterior partial slices are removed. We then remove every other slice so that the thickness is 50 µm similar to the thickness of an experimental dataset which is usually 40 µm to 60 µm. For each slice in the simulated brains, we apply a random rotation that is smaller than 2.5°clockwise or counterclockwise to simulate the in-plane rotation, because our experimental slices are not always straight up, and a translation smaller than 10 pixels in each direction. We then randomly pick a nonrigid transformation which was generated when we used our MRF nonrigid registration method based on salient points to map a real experimental slice to its corresponding atlas slice.

The most common metric for evaluating image registration is the target registration error (TRE) measured as the Euclidean distance between landmark point coordinates in the moving image mapped by a computed transformation to the fixed image and the corresponding landmark points in the fixed image. For the evaluation of sectional brain registration, we asked one of our neuroanatomist coauthors in the study [53] to identify 20 sparsely-scattered landmarks in the corresponding sectional atlas which she would be confident in locating in both simulated and real experimental brains regardless of their quality. In the full brain, we select 17 regions - 81 lateral ventricle, 581 triangular nucleus of septum, 286 suprachiasmatic nucleus, 338 subfornical organ, 223 arcuate hypothalamic nucleus, 830 dorsomedial nucleus of the hypothalamus, 470 subthalamic nucleus, 884 amygdalar capsule, 587 nucleus of darkschewitsch, 214 red nucleus, 931 pontine gray, 872 dorsal nucleus raphe,

642 nucleus of the trapezoid body, 574 tegmental reticular nucleus, 169 nucleus prepositus, 222 nucleus raphe obsurus, 207 area postrem (the numbers in front of region names are the region ID in the annotation volume) and generate landmarks automatically by sampling 100 points along these brain region boundaries. These regions show contrast to their neighboring regions in at least a few slices that contain them. The sampled points cover 32% of the total number of slices and 72% of the entire brain length. We then map the selected landmark points with the known transformation to obtain the ground truth.

In addition to TRE, in the sectional simulated brain, we can compute the true error of both our method and of the expert, since we have ground truth. Because TRE measures the distance between expert marked landmarks and the new landmark locations after we warp the moving images with our method, it represents a combined error of expert error and computation error. Getting the computation error and expert error separately on simulated datasets will help us interpret the TRE we measure on real experimental datasets. For the full simulated brain, we are only able to measure the computation error since the landmark points are generated automatically, but the expert error from the simulated sectional brain experiment can help us interpret this result.

We choose to compare our method with a reconstruction-first method, because the main difference between our method and previous methods is that our method does not require a prior reconstruction at all. As reviewed in Section 2.2, there were many previous methods to choose from. Because most experimental datasets contain many distorted slices, we wanted to choose a reconstruction method that can nonrigidly correct these distortions. Second, we avoided methods that required much manual work. These include reconstruction methods that require a selection of reference slice [7, 48] or blockface image [19] and frameworks that do partial reconstruction and solve the mapping problem iteratively because the selection of the first reconstruted section is highly variable [25, 68]. Third, we wanted a method that would work on different brain sections and the full mouse brain. Because Ju *et al.* 's method [32] provides an automatic program that reconstructs a brain volume nonrigidly and does not need any manual selection, we used their tool to reconstruct the moving brain. We first rigidly aligned simulated slices from the middle slice to the two ends, then nonrigidly reconstruct the sections with a five-slice neighborhood through Ju's elasticAligner [32]. With elasticAligner, warp functions between every pair of slices were computed. The weighted average of these warps in the five-slice neighborhood was then applied to every slice. Ju *et al.* 's method only reconstructs a brain volume. We need to choose another method to register the reconstructed volume to the reference volume. We tested different methods and finally chose 3D registration with Elastix [34, 56] using the parameter file created by Hammelrath *et al.* [27]. This parameter file was developed specifically for 3D mouse brain registration and performs the best among the files that we tested. Reconstructed volumes were first rigidly aligned, then affinely aligned, and finally elastically aligned with B-spline to the reference volume. To have a fair comparison, for the sectional simulated data, we use the same exact reference section in the atlas for both our method and the comparison method.

### 3.7.1 Quantitative Result

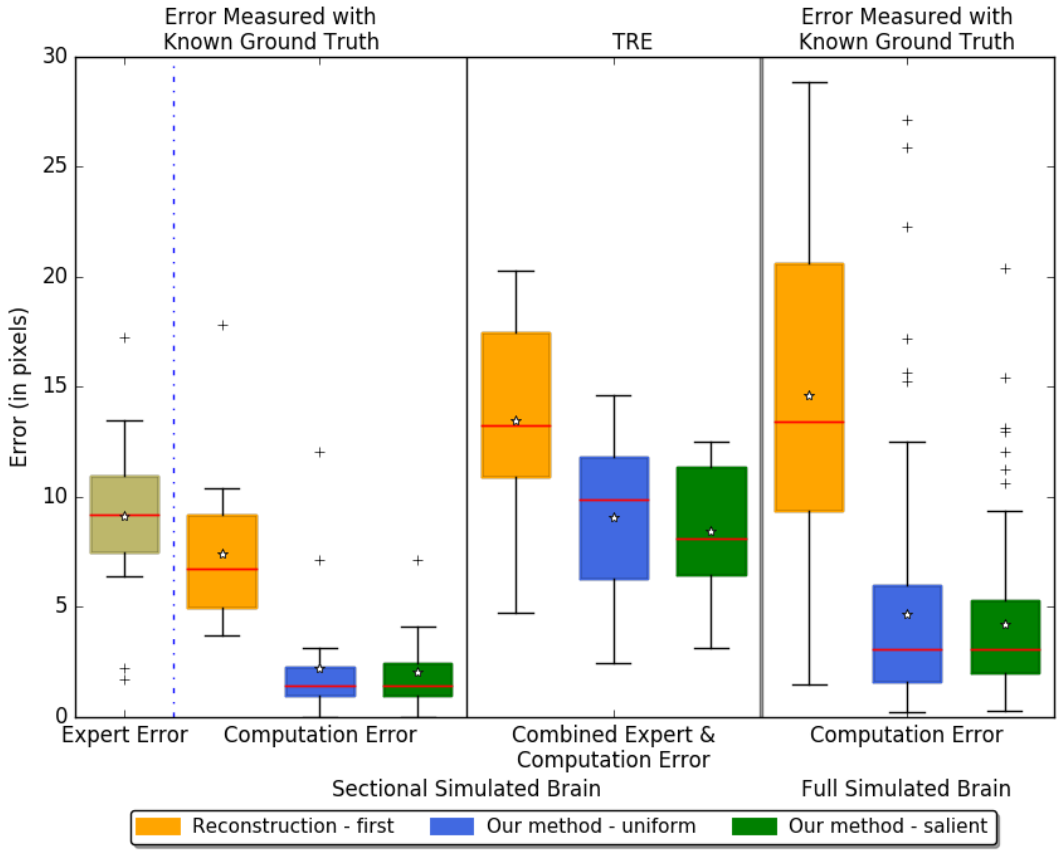



Figure 3.10: Boxplots showing results on the sectional simulated brain and the full simulated brain. The boxplots in the left two columns show the sectional brain results. We measured the intrinsic expert error, pure computation error of the reconstruction-first method, our method with the uniform grid in the 2D non-rigid registration part, and our method with the salient control point in the 2D non-rigid registration part, and the TRE — combined expert error and computation error. The third column shows the results on the full simulated brain where we measured the pure computation error only. The lines on the boxes represent the minimum, first quartile, median (red), third quartile, and maximum respectively. The star denotes the average. The crosses mark the outliers that are not within 1.5 interquatile range.

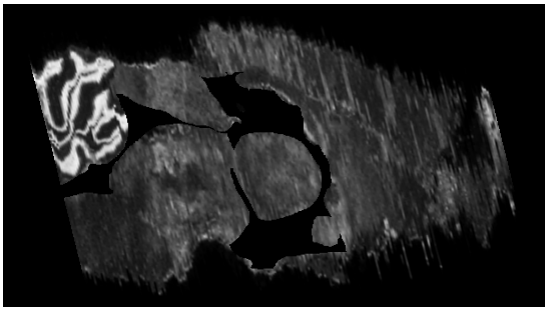
Figure 3.10 reports the results of our simulated datasets — separately measured expert error, computation error, and the TRE (combined expert and computation error) of the reconstruction-first method and our method. We evaluated our method both with the uniform control points and the salient control points, as shown in the blue and green boxes. Because all the terms in our energy function are normalized to be within range $0-1$, we did not heavily tune the parameters to obtain these results. We only tuned the coefficients before each term in our energy function with several

images to get a good visual result. The expert has an intrinsic error of about 9 pixels (one pixel equals 25 µm as given by the resolution of the atlas volume of the ABA — similar to the TRE of our method on the sectional simulated brain. This figure also shows that our methods, with a 2.24-pixel and a 2.07-pixel error on the sectional simulated brain, is about three times better than the reconstruction-first approach which has an error of 7.45 pixels. For the full simulated brain, the TREs of our method are 4.68 pixels and 4.24 pixels vs. the TRE of the reconstruction-first method — 14.59 pixels. The computation errors are very similar between our method with uniform grid and with salient control points as expected, because using salient control points aims at improving smoothness in the regions where intensity values are very similar. We expect the computation error to be slightly greater and show more variance for the full simulated brain than the sectional brain, because even though the landmarks are sampled on region boundaries that show contrast to neighboring regions at least in some slices, we did not further constrain on the slice numbers when we sampled the landmarks. Therefore the landmark points may still fall in a region that has relatively uniform intensity which makes the corresponding location of these points ambiguous. Moreover, because the simulated deformations are taken from real image pairs, when applied on a slice in the full brain, some large deformations unavoidably are applied to regions with relatively uniform intensity. Without features salient enough as in the original image pair where the simulated deformation comes from, it is difficult to fully correct the distortions. The results agree with our expectation, and the error of our method is still much smaller than expert error on the sectional simulated brain and the reconstruction-first method error. Results from both the sectional simulated brain and the full simulated brain show that using salient control points generates less outliers and basically means the salient control point method generate smoother and better warps.

### 3.7.2 Qualitative Result

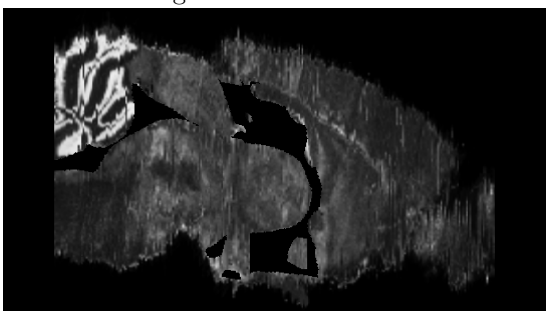
The qualitative result of the comparison experiment is shown in Figure 4.3. It displays the sagittal view of the moving volume, the reference volume, and the overlaid image for each method. We masked out the ventricle systems or the empty space in the brain volumes. As one can see in the second row, the sagittal views of the atlas volume are different. It is because the comparison method reconstructs the moving volume and registers the volume to the reference volume directly, while our method finds the best cutting angle and reorient the atlas so that the atlas is sectioned in the same angle as our moving volume. Therefore (D) shows the sagittal view of the rotated atlas while (C) shows the sagittal view of the original atlas. Another difference is that because our method does not reconstruct the moving volume at all, (B) is generated for the illustration purpose only. With our method, one would be able to get mapping from every pixel in the moving image stack to the corresponding voxel in the reference atlas volume. To show the "reconstructed" brain with our method, we place each slice to the coordinates of the atlas volume rotated with the best cutting angles and interpolate the volume in the anterior-posterior direction to fill in "missing" slices, because our simulated slice is

Reconstruction-first

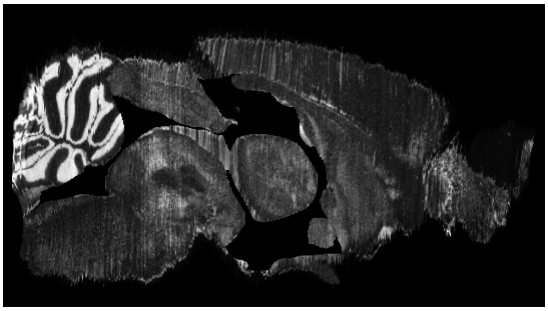


(A) Reconstructed simulated volume registered to the atlas volume

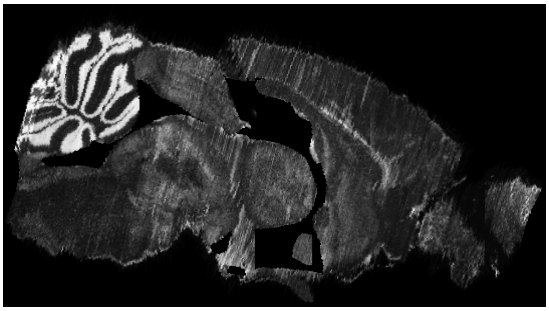
Our method - general



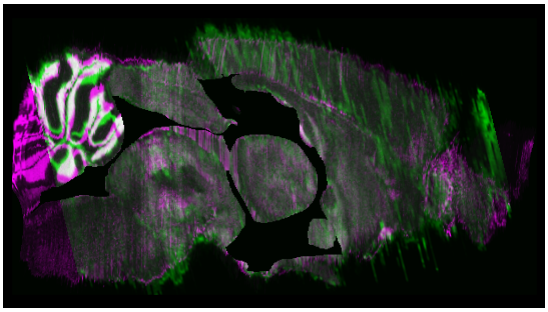
(B) Interpolated simulated volume by placing each slice to the coordinates of the rotated atlas



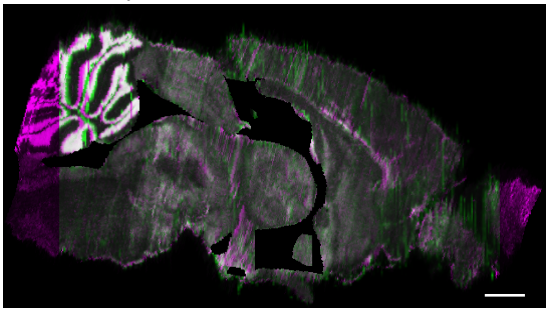
(C) Original unrotated atlas



(D) Atlas rotated with the best cutting angles returned by our method



(E) Overlay of A (in green) and C (in magenta)



(F) Overlay of B (in green) and D (in magenta)

Figure 3.11: Full simulated brain results, sagittal view. Scale, 1 mm.

"thicker" than the atlas. For the reconstruction-first method, we can show the reconstructed volume directly. As one can see in (A), the reconstructed volume is very smooth. The last row shows the overlaid of the moving volume and the reference volume after registration. The comparison method seems to register the central section with the atlas well. But it is clear that the front and back of the brain are misaligned. Because the goal of our method is to map the experimental slices to its corresponding coordinates in the atlas volume, we emphasize more on the correctness of alignment rather than smoothness. As shown in (F), with our method, the warped moving volume is positioned correctly on top of the atlas volume. Because the moving volume is shown in green, and the reference volume is shown in magenta, a perfect alignment of intensities would result in white. The merged image in (E) is greener in color meaning the error of the comparison method is also greater than that of our method.

With this comparison experiment between our method and a reconstruction-first method on the simulated data, we demonstrated that our method gives less TRE because our method does not have a reconstruction step at all avoiding the z-shift problem. By showing that our method has better TRE than the comparison method on the sectional simulated brain, we also proved that even if a small brain section is reconstructed, a small error will still be introduced.

# Chapter 4

# Applications

In the previous section, we have described our method and its performance on simulated datasets. In this section, we talk about the application of our method on various real experimental datasets. We hope these examples provide potential users a better idea of how to use our method to register their data even if the new datasets are different than the experimented ones presented.

Each lab usually has their own image preprocessing pipeline that reads the digital image files from their microscopes. These preprocessing pipeline usually crops individual slices out from a large scan containing multiple adjacent slices, orders them, normalizes the intensity to 8-bit or 16-bit, and downsample the images if needed. Our pipeline read in preprocessed and ordered images in the TIF format with additional requirements listed here:

1. If tissues are generated with free-floating immunohistochemistry, the flipped image should be corrected to reflect the image viewing from the front of the tissue. The more left-right asymmetry there is in the image file, the more importance this step is.

2. Depending on the lab's image processing pipeline, some experimental datasets will have excessive background (nonzero-intensity regions that does not represent real tissue). This background noise needs to be removed for our image registration pipeline to work correctly, because we use the shape information to perform the initial affine registration in both the plane-wise mapping and pixel-wise mapping steps. Empty space inside real tissues does not need to be removed.

Some brain sections consist of loosely-connected tissues. These tissues may be lost or displaced during the processing steps. Our method works well when an experimental brain or a sectional experimental brain contain some of these slices. However if these slices form a large portion of an experimental dataset, additional preprocessing will be needed. We will discuss the needed processing later in this chapter.

## 4.1 Nissl-stained Full Brain Data

Our first experimental dataset was a full brain dataset, Nissl stained, 50 µm thick, consisting of 202 slices. Since this was the first experimental dataset that we worked on, we manually selected 35 good-quality slices that were relatively less distorted comparing to other slices as the Set A which was used to compute the best cutting angle. The average index difference between the best matching upper and lower half slices are shown in Figure 4.1, which shows the expected pattern. The average index difference is negative for a cutting angle that is smaller than the actual cutting angle and is positive for a cutting angle that is greater. The best cutting angle is at where the average index flips sign and gives the smaller index difference. According to the chart, 11 is the best cutting angle.

We found the best cutting angle around the superior-interior axis similarly with the left and right half slices. The angle around the superior-interior axis is zero. In later reviewing this data, we noticed clear left-right asymmetry which would indicate a rotation. This discrepancy was tracked to the slice preparation using free-floating methods, which flipped some slices back side up. Since the images were only preprocessed to make sure the ordering is right, and the left-right flip was not corrected. Therefore our algorithm detected no rotation around the superior-anterior axis. We attempted to correct this error after the problem was identified, however since the number of potentially "flipped" slices were very close to the "non-flipped" slices and we were not neuroscientist, we did not fix the problem. If an application requires high precision of the mapping, and images are generated free-floating, making sure images are front-side-up is very important.

After the best cutting angles were found, we found the plane-wise mapping by running the same algorithm on the full selected slices and interpolated the array to find the correspondences of the slices which were not in the selected subset, A. We sampled every $20^{th}$ slice in this experimental mouse brain and showed their corresponding resectioned atlas plane detected by our algorithm in Figure 4.2. Slices are ordered from posterior to anterior. Odd rows are the experimental slices, and the corresponding atlas slice is shown directly below. From the image we can see clearly the perfusion and hydration process caused enlarged ventricle in the experimental images. There are also many slices that are corrupted, especially towards the anterior or the posterior brain. None the less, our algorithm could detect very closely matching atlas image for the experimental slices even for the ones with artifacts. We can see that the most anterior slices are not matched as well as other slices, because many anterior slices are too small and lost during the imaging procedure. We had a neuroscientist check our result. He said that the result was better than his manually identified result because computer algorithm could take into account different cutting angles and also spatial constraint at the same time, and humans were bad at computing 3D projection in mind. We did not further evaluate the result of the full experimental brain, because the data was not linked to any study.

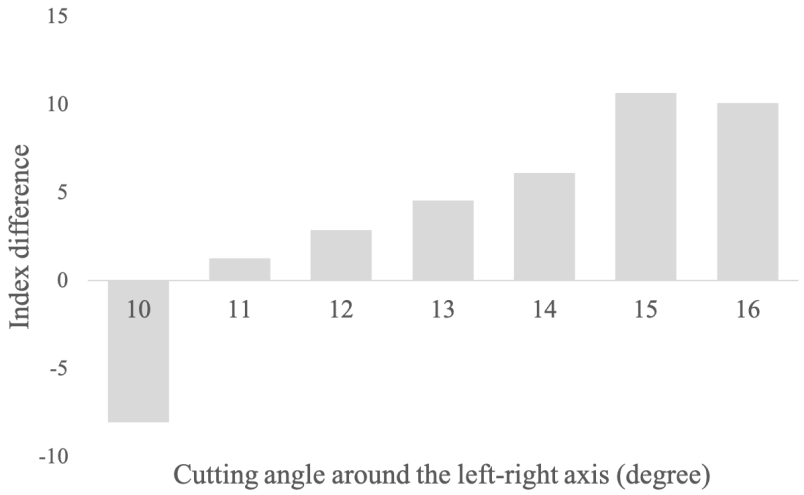



Figure 4.1: Average index difference between the upper and lower best matching slices of selected experimental images.

## 4.2 Dorsal Raphe Sectional Brain Dataset

The first real application of our registration framework was a systematic anatomical study in the hindbrain to study the organization of the dorsal raphe (DR) serotonin system and its behavioral functions related to depression and anxiety [53]. Prior this study, most of the research community had viewed the DR serotonin neurons as a single distribution. The authors of this paper were interested in further studying the distribution of serotonin neurons projecting to different brain regions in the DR region. These serotonin neurons were stained with anti TPH2 antibody and retrograde traced with viral-genetic methods. Examining them requires sectioning the DR brain section into thin slices. Obtaining the distribution requires mapping these neurons from different brains in the same experiment to the atlas coordinate. Because it is hard to section experimental brains with the same angle as the atlas, and the DR is a very small brain region, in order to draw useful conclusion, a precise mapping algorithm was needed.

### 4.2.1 Data Description

There were 36 Nissl-stained sectional brains in this study. Each sectional brain consists of 30 to 55 coronal slices sectioned with a thickness of either 40 µm or 50 µm. The resolution in the sectioning plane was measured to be 5.1 µm per pixel. Image size varies across brains with resolution ranging from 1 megapixels ($1000 \times 1000$) to 6 megapixels ($2000 \times 3000$) in the sectioning plane depending on the size of the original brain and the imaging space. Some brains were processed with the free-floating method, and some were processed with the slides-attached method. Our program mapped all the sectioned brains to the atlas, and with the obtained transformation fields, we were able to estimate the

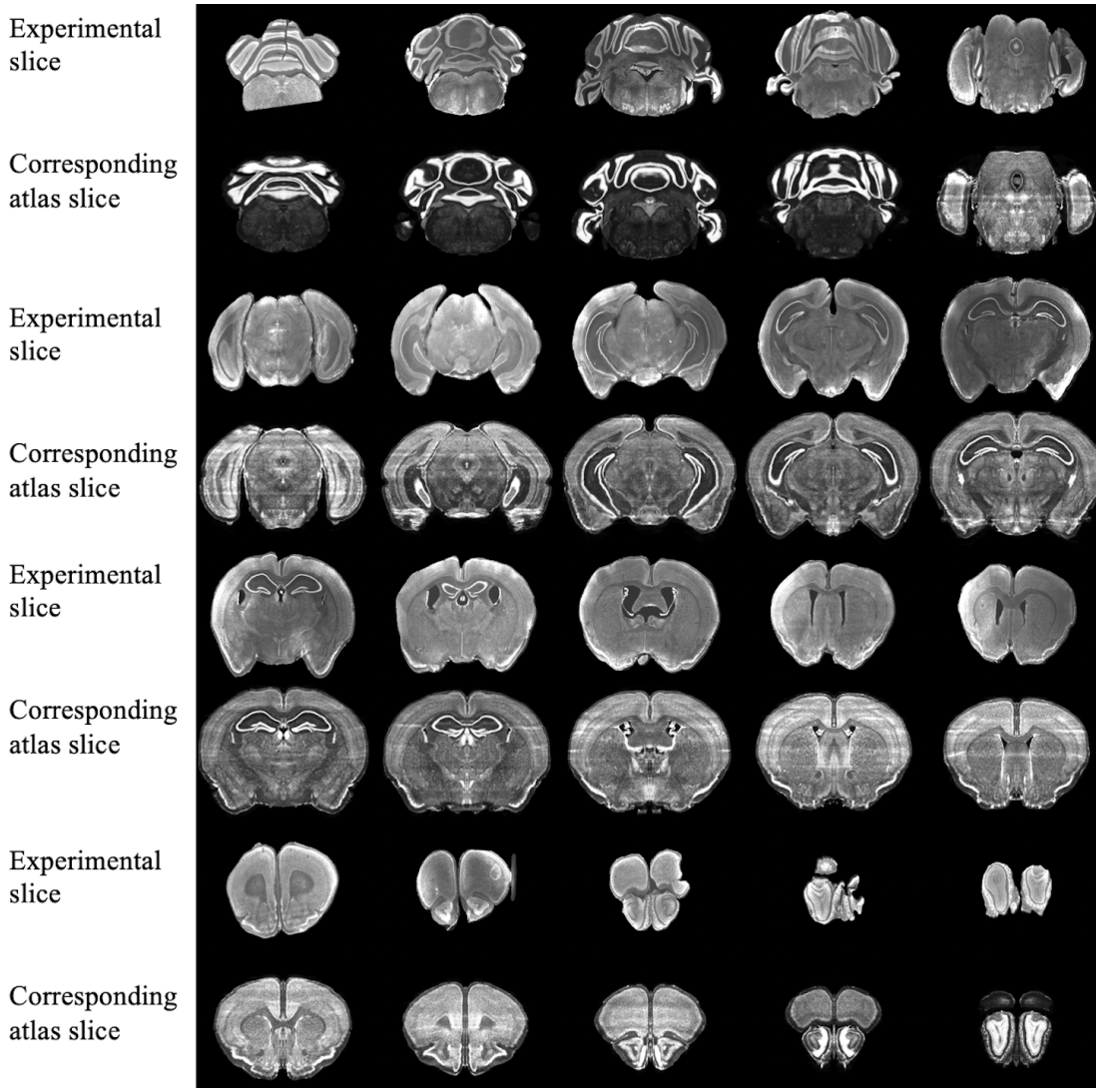
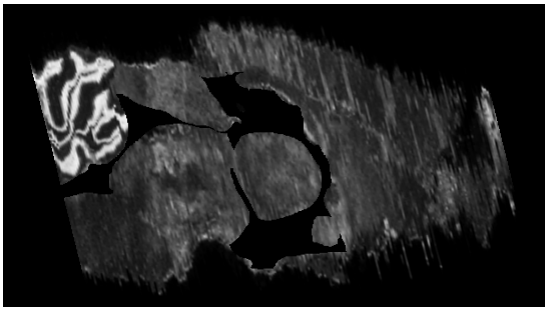


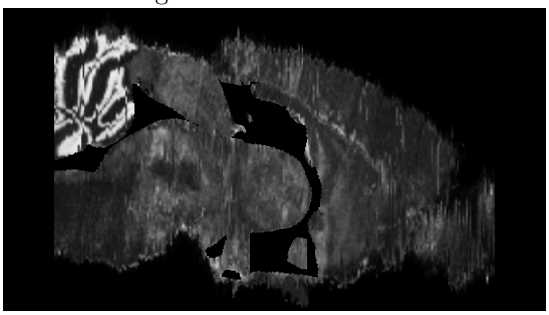

Figure 4.2: Sample experimental slices in the full brain dataset and their corresponding plane computed by our algorithm.

Reconstruction-first

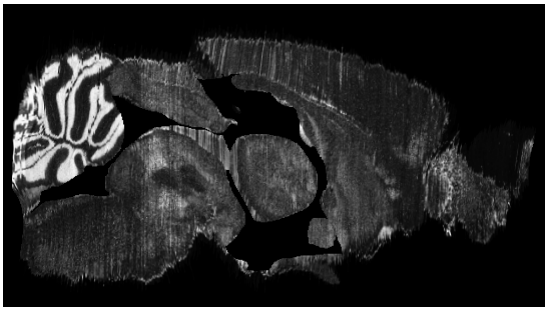


(A) Reconstructed simulated volume registered to the atlas volume

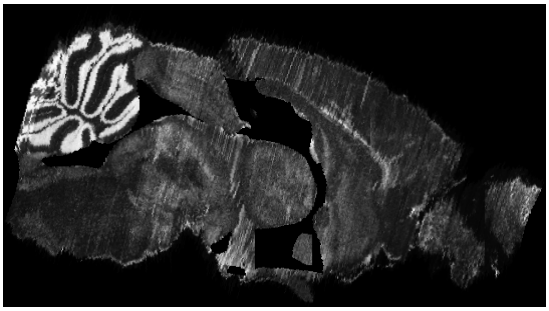
Our method-general



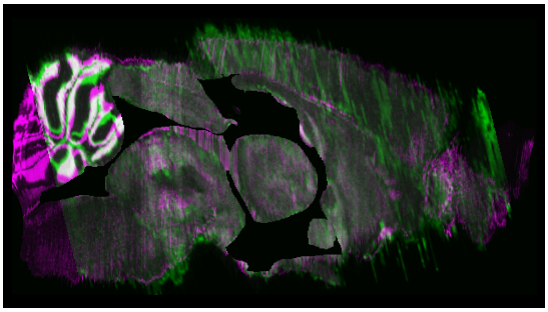
(B) Interpolated simulated volume by placing each slice to the coordinates of the rotated atlas



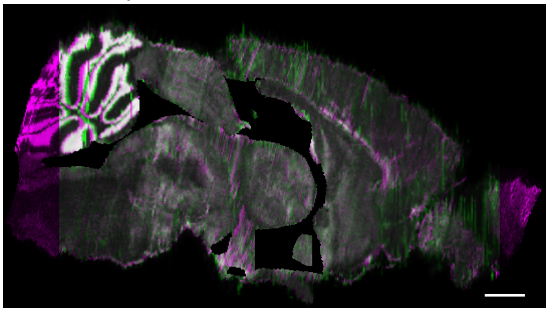
(C) Original unrotated atlas



(D) Atlas rotated with the best cutting angles returned by our method



(E) Overlay of A (in green) and C (in magenta)



(F) Overlay of B (in green) and D (in magenta)

Figure 4.3: Full simulated brain results, sagittal view. Scale, 1 mm.

model for serotonin neurons localized with different distribution projecting to different downstream brain regions that relate to distinguish behaviors.

### 4.2.2 Preprocessing and Creation of the Sectional Atlas

The ideal case in a biomedical registration problem is that the moving image contains the same structures as the fixed image. As described in Section 3.1, our HOG similarity metric can measure similarity between two images even with the presence of distortions. The side-effect is that the metric is not very sensitive, and finding best corresponding slice requires additional constraints. The posterior cerebral cortex is easily detached during the slice preparation procedures. It presented a problem when we tried to find the best matching slice of a slice that had misplaced or missing tissue, because we aligned the slice with a resectioned atlas slice affinely based on the same information before computing the HOG difference.

Therefore we needed to preprocess both the atlas and the experimental dataset so that the two datasets contain similar information. We made sure all the easily-missing tissues were removed in both experimental and the reference atlas. Because most sectional brains in this study only consisted of 30 slices, and there were not as many salient features in this brain section, we wanted to keep as much information as possible. Therefore based on how much the easily-detached tissue were still attached, we grouped the 36 experimental brains into three groups — one with all the easily-detached tissues removed (some brains processed free-floating), one without cortex but with cerebellum (the remaining brains processed free-floating but had more tissues attached), and one with all the tissues (brains made slides-attached). We created a sectional atlas according to how much tissues were kept in each group. We split the brains made with free-floating method based on their quality. Because most of our sectional brains consisted of only 30 slices, we wanted to keep as much information as possible.

To make the sectional atlas, we first chose a slice index range in the unrotated atlas that covered all the regions that may appear in any of our experimental sectional brains, then we removed easily-detached tissues in that group. We made use of the annotation volume of the ABA. Because the annotation volume was not very accurate, our neuroscientist co-author further corrected some removed regions to make sure all target tissues were correctly removed. For the experimental brain sections, our collaborators removed the easily deformed regions based on which group the sectional brain belonged to. This step was completed during the preprocessing steps that generated clean and ordered image stacks.

Even though the experimental slices were preprocessed by operators to remove the nonzero-intensity background and to keep only the tissues, this procedure was not quality-controlled. It would cause problems when this preprocessing was not done well. The “fake” gradient from these nonzero intensity background pixels to the zero-intensity background affected our results. If excessive background still remained around the real tissue, the outer contour of images might not align well,

because our cost function consisted of a symmetric difference term. Therefore we resegmented the experimental images again before registering them to the reference. We used a similar method to preprocess the atlas images as in Section 3.3.3, except that the masks were morphologically eroded and dilated with a disk of 3 pixels, and a computed background region would be kept if its area was greater than 50 pixels. These parameters were selected based on experiments on several slices in one of our experimental brains. Same parameters were used for all slices from all brains. In the case that a slice was missing a relatively large portion of tissue, after the plane correspondence was found, a manual tissue removal would be done on the fixed image to crop out the same corresponding portion that was missing before the 2D nonrigid registration.

### 4.2.3 Selection of Set A

In Section 3.6, we briefly discussed the general guideline of the selection of Set A. In this anatomical study, most sectional brain consisted of only around 30 slices. We used every third image for most of the brains — about 10 slices for each experimental brain. For some brains with relatively more damaged slices, we manually checked the automatic selection and replaced slices with significant damage with a nearby good quality slice. With Matlab, it takes 38.8s on a 12-core 3GHz Linux machine to evaluate a set of 12 slices, or equivalently to evaluate a cutting angle on a sectional brain.

### 4.2.4 Improvement based on Data-specific Properties in 2D Non-rigid Registration

The dorsal raphe nuclei are ventral to a hollow structure called aqueduct. Due to the difference in brain preparation procedure, for example, the dehydration step, as illustrated in Figure 4.4, the aqueduct showed variability between our experimental brains and the atlas. The difference in size, appearance, and especially edge orientation of the aqueduct made aligning the regions around it difficult using the HOG descriptor alone.

The solution was to force a correspondence of this region between the moving and fixed images, which it required a segmentation of the aqueduct in the experimental images first. Because aqueduct appearance varies across subjects, sectioning angles, and longitudinal axis as shown in Figure 4.4, it is hard to segment them with a single traditional segmentation method. Therefore we used a learning-based approach which we will talk about next.

#### 4.2.4.1 Learning-based Region Segmentation

**Data Augmentation** Because of the aqueduct variability we saw in our experimental datasets, an unsupervised-learning approach would not work. To use a supervised learning approach, we needed to address the lack of labeled data. Our experimental dataset consisted of 36 brains where most brains have about 30 slices. One slice would have at most one aqueduct region in it. The total amount of

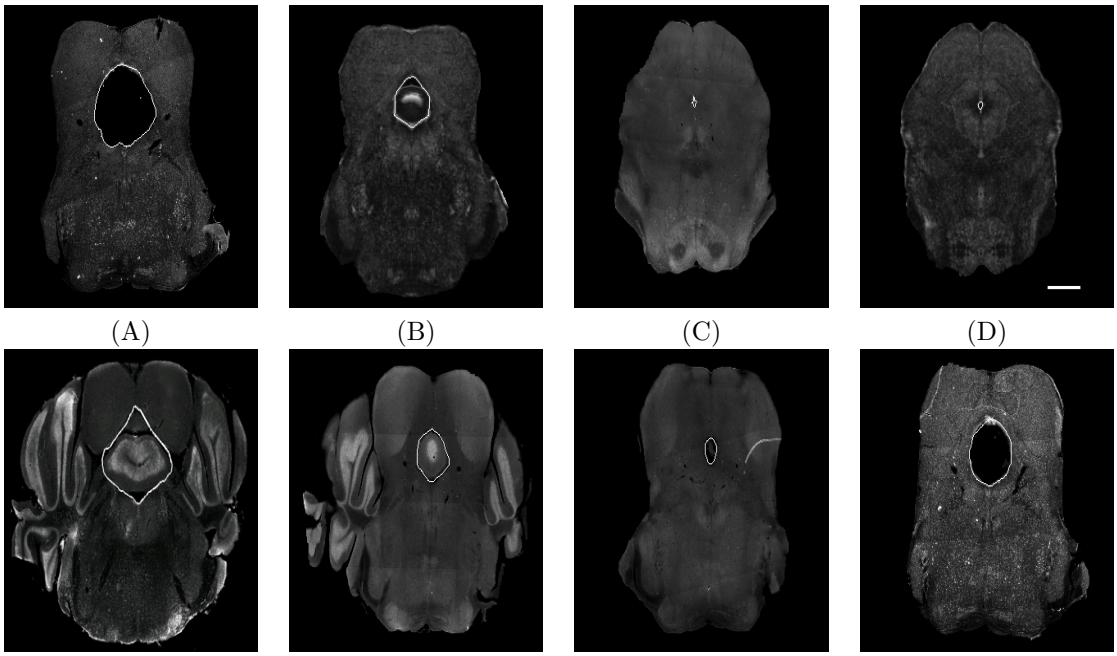



Figure 4.4: (A) Expanded aqueduct in an experimental image. (B) Aqueduct in the corresponding atlas image of (A). (C) Squeezed aqueduct in an experimental image. (D) Aqueduct in the corresponding atlas image of (C). (E)—(H) Various aqueduct appearance in different brains and slices. Aqueduct contours are marked with white curve. Scale, 1 mm.

| Layer | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Convolution | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $1 \times 1$ |
| Dilation | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 1 |
| Truncation | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Receptive field | $3 \times 3$ | $5 \times 5$ | $9 \times 9$ | $17 \times 17$ | $33 \times 33$ | $65 \times 65$ | $129 \times 129$ | $257 \times 257$ | $257 \times 257$ |
| Output Channels | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | C |

Table 4.1: Aqueduct segmentation network architecture. C is the number of channels in the input image.

images we had was limited, and we could not ask our neuroscientist coauthor to manually segment all images. Therefore we randomly selected five brains as the initial dataset to generate the training data and asked the neuroscientist to annotate the aqueduct in all the 150 images. Then we did heavy data augmentation on this initial dataset. We applied randomly selected affine transformation, then randomly selected nonrigid transformation to each image and the annotation mask. We then resized them to $512 \times 512$. A left-right flip is further applied to generate more training data. In the end, more than 5000 images and annotations were generated.

**Neural Network** Despite the heavy data augmentation, our training set was still small, so we prioritized networks with a relatively small number of weights and larger pretrained networks where we used transfer learning, and only trained our labeled datasets over the last few layers. We first found the transfer learning approach did not work very well. Our guess is that those networks were previously trained very well on natural images, and since our dataset is very different, the outcome depends heavily on in what layer of the original network we started to train our data. Among the relatively smaller networks, we finally selected a network similar to Chen *et al.* 's [14] implementation of context aggregation [69], a network built with mainly dilated convolution layers. The parameter chart is shown in Table 4.1. Because the aqueduct would only appear at a relatively stable location in a slice, it was not translation-invariant. Unlike the translation-invariant features where we only need the receptive field of a filter to cover the size of the object, we need to grow the receptive field of the layers to cover almost the entire slice so that we can "code" this relative location information in the slice into the network. The cascaded dilated convolutional filters work very well for our limited data because the number of parameters grow linearly while the receptive field grows exponentially. We use intersection over union (IoU) as the loss function, and leaky ReLU as the activation function.

We trained the network until the IoU stopped increasing. We ran the network on the remaining brains. Partially because the randomly selected brains that was used to generated the training dataset do not represent the full variability of all experimental brains, about 10% of slices need a manual segmentation.

#### 4.2.4.2 Integrating Region Information

We now have the contours of the aqueduct in both experimental and reference images. Because the shapes are sometimes much different, we could not use shape context as what we did in Section 3.1 to build point correspondence. Contour alignment methods such as minimizing the Hausdorff distance are computation-expensive. Therefore we used a simple way to match the aqueduct contour points. Because the images are straight-up, we found the highest and lowest points on the aqueduct contours and used them as the corresponding points. If there were multiple highest or lowest points, we chose the point that is closer to the centroid of the aqueduct contour. We then divided the contour curves into the left and right halves with the highest and lowest points and built more point correspondence by uniformly sampling eight points along each half.

We nonrigidly registered the image pairs in this study with the uniform-grid MRF model described in Section 3.3. These control points from uniform grid would not have enough flexibility to align the aqueducts when the aqueducts were smaller than the grid spacing. Therefore we used these corresponding points on the aqueduct contours as an additional term in the energy function. The term measured the Euclidean distance between the warped experimental aqueduct contour points and their corresponding atlas aqueduct contour points:

$$D_p^t(l_p) = d(\mathbf{v_p}, \mathcal{T}_p^t(\mathbf{u_p}, p)) \tag{4.1}$$

where an experimental aqueduct contour point $u \in \mathbf{u_p}$, if its influence to node p - $\eta_{t+1}(|u-p|) > 0$. $\mathbf{v_p}$ are the corresponding aqueduct contour points in the atlas image, and $d$ measures the Euclidean distance between two sets of points. The energy function on coherent nodes now becomes:

$$E_{coherent}^t = \sum_{p \in \mathcal{G}_{coherent}^t} \mathcal{F} \circ V_{p,atlas}^t(l_p) + \sum_{p \in \mathcal{G}_{coherent}^t} \mathcal{F} \circ D_p^t(l_p) + \sum_{p \in \mathcal{G}_{coherent}^t} \sum_{q \in \{q : e_{pq} \in \mathbf{E}^t\}} \mathcal{F} \circ V_{pq}^t(l_p, l_q) \tag{4.2}$$

Solving this energy function could only register the aqueduct region well if the size of the aqueduct is on the order of, or larger than, the control point spacing. Registration is difficult if the aqueduct is squeezed and is much smaller than the spacing. Therefore we further warp the corresponding points on the aqueduct contours with thin plate spline (TPS) [10], and because our other tissue regions are mostly registered, we add surrounding control point locations into the TPS as well so that we do not sacrifice other registered regions.

We improved the 2D nonrigid registration algorithms in Section 3.4 by using better located control points than points on a uniform grid. In this case, we may included some the aqueduct points as our control points. The registration would work well if the aqueduct is on the order of at least half of the spacing, and the aqueduct has similar HOG profile. To make sure the aqueduct was registered well in all our images, we also could added all these sampled aqueduct points first as our control points,

and then detect other control points. We experimented with the salient control points method both without adding the sampled aqueduct contour points and with those points. Like in the uniform grid formulation, aqueduct control points and the salient control points whose patch would cover some of the aqueduct correspond to the distance term $D_p$, and the salient control points correspond to the HOG similarity term $V_p$. If a control point is on the tissue boundary, a symmetric difference term is also associated with it to better align the image contours. The energy function will be:

$$E^t = \sum_{p_{salient}} \mathcal{F} \circ V^t_{p,atlas}(l_p) + \sum_{p_{aqueduct}} \mathcal{F} \circ D^t_p(l_p) + \sum_p \mathcal{F} \circ S^t_{p,atlas}(l_p) + \sum_p \sum_{q \in \{q : e_{pq} \in \mathbf{E}^t\}} \mathcal{F} \circ V^t_{pq}(l_p, l_q) \tag{4.3}$$

where $S$ represents the symmetric difference term; control points $p$ is grouped; and a regularization term is added regardless of point $p$ and point $q$'s category. Because the aqueduct boundary points are directly the control points, we do not need to add a post TPS deformation like we did with the uniform grid approach.

### 4.2.5 Neuron Mapping

Each retrogradely labeled cells in the DR is imaged again with confocal imaging with higher resolution than with the fluorescent microscope to record the positions of retrogradely labeled cells from different projecting sites. A merged channel confocal image of the DR region is shown in Figure 4.5. Individual cells can be seen clearly, and we can split channels to see neurons from different projecting sites.

We registered the full Nissl-stained coronal section to the atlas and obtained first a sectioning plane in the atlas and then a nonrigid in-plane mapping to the corresponding atlas plane. The confocal images contain more precise location information of the labeled neuron positions and their original injection site information. Despite being taken from the same physical slice, the coordinate information from these two different imaging types was different — the resolutions were different, and the confocal images were just a "crop" centered at the dorsal raphe region. Therefore the obtained transformation did not directly transfer. Since the DR region is directly under the aqueduct (aqueduct contour labeled in white in Figure 4.5), we used the lowest point of the aqueduct as a reference point and measured each neuron" distance in the horizontal and vertical direction relative to this stable feature. With the resolution of confocal image as well as our full histological image we could compute the distance of these neurons to the reference point in pixels.

The problem became computing the position of the reference point in the experimental image. We directly took from the annotation the lowest aqueduct point $C'$ in the atlas and found the corresponding position in the experimental image with Equation 3.23. The absolute position of labeled neurons in the experimental image are then computed with the position of the reference point and the horizontal and vertical offset. After we get these positions, we use the above equation again

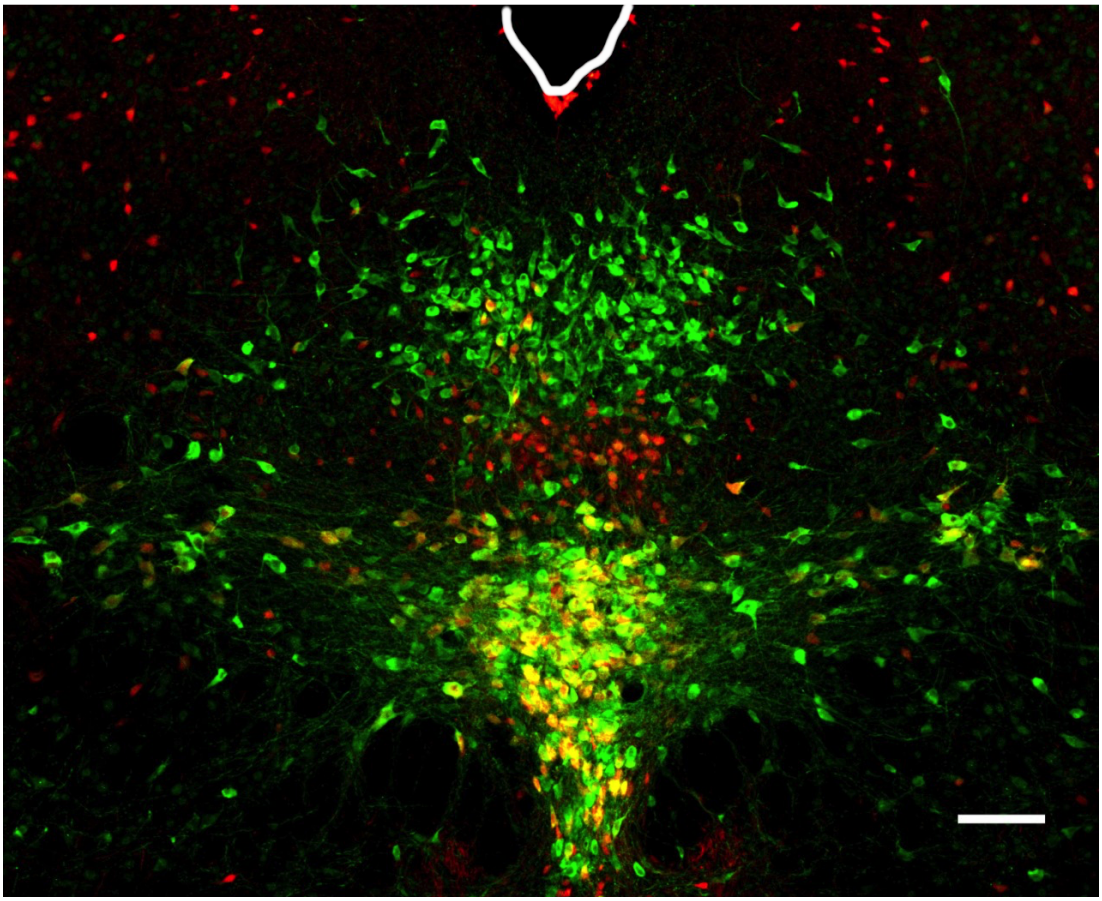


Figure 4.5: Example coronal confocal section from the DR study. Different color represents different injection sites. Splitting channels reveals position of neurons from a single injection site. White line, aqueduct border. Scale, 100 µm. [53]

to find the corresponding position in the atlas.

### 4.2.6 Evaluation

We used five brains in this anatomical study to analyze the performance of our framework on real experimental data. We used the same comparison method as described in Section 3.7. In our paper [66], we had not implemented the salient control point based nonrigid registration, therefore we compared the performance of our method with uniform grid without and with the aqueduct segmentation. To evaluate the performance of the salient control point 2D nonrigid registration method on real experimental data, we evaluated it on the real data from the DR study. With the evaluation results from [66], we added the new results and plotted them together in the boxplot in Figure 4.6. It reports results of the reconstruction-first method, our method without and with the data-specific improvement that deals with the aqueduct registration, and the method with salient points as control points.

The five brains represent some of the data variability we saw in the real datasets in this study. These TRE numbers shown in the chart represent a combination of expert and computation error because it measures the distance between corresponding points returned by our methods and the points located by a neuroscientist. Based on the simulated result on the sectional simulated brain in Section 3.7, we believe the intrinsic expert error is likely to be much larger than the computation error of our methods. With the data-specific improvements, even though the average TRE did not improve much, the standard deviation of errors became much smaller. As expected, we see a larger improvement on landmark points near the aqueduct. The TRE ratio is on average 2.56 between the reconstruction-first method and our methods.

The additional experiment we did is given in the green boxes - evaluation of the MRF nonrigid registration with detected salient control points. Using salient control points improves the average TRE from using uniform grids by 10.9% on average and reduces the standard deviation by 18.9% on average. Comparing to using the uniform grid and detecting and warping the aqueduct regions separately, using the salient control points instead shows both relative better and worse results. It is because with the salient point control, we use spacing that is comparable to the uniform grid spacing, and the aqueduct size is sometimes smaller than the spacing. The results show salient control points gives better performance than using the uniform grid, but improvement based on data-specific properties is still necessary to achieve the best performance. To explore further, we added the sampled aqueduct contour points as salient control points, and then detected other salient control points. Adding these points improved results from the general salient control points approach. For the aqueduct contour points, we computed both the HOG similarity term and the distance term and the symmetric difference term and weighted them so that each control points would contribute equally.

Because our experimental sectional brains only consist of about 1/7 to 1/6 of a full brain length,

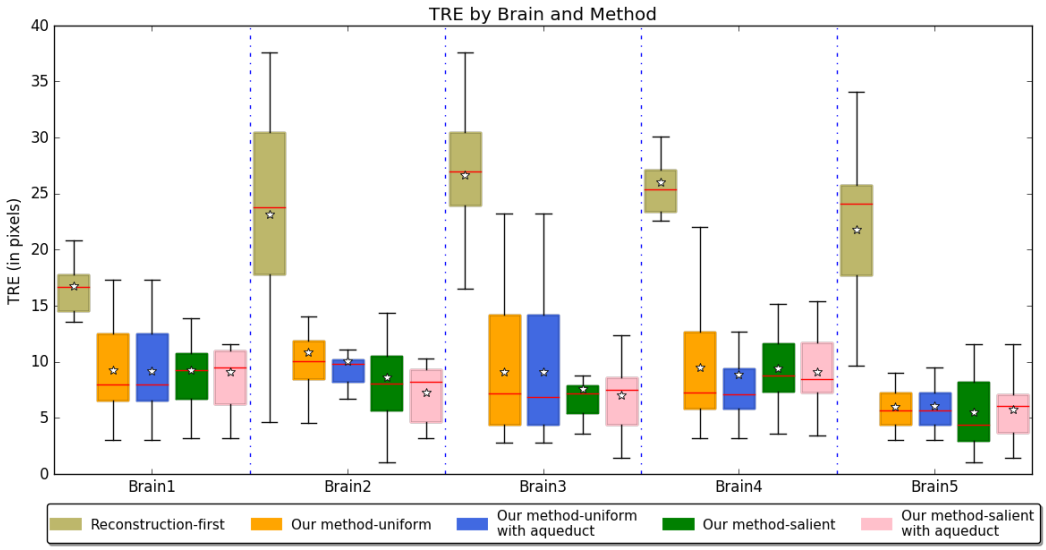



Figure 4.6: Boxplots of the TRE on evaluated experimental brains.

showing the sagittal view of these thin stacks does not exhibit the correctness of alignment. Therefore, we show four evenly-spaced slices in each experimental brain and their corresponding planes in the atlas volume after we mapped them to the same coordinates. Figure 4.7 displays a triplet of images for each slice location and for each experimental brain. Each row shows two triplets of images. In each triplet, the left image is the registered slice using our method, the center is the corresponding atlas plane, and the right image is the registered slice using the reconstruction-first method. To give the registration-first method a fairer comparison, we try to avoid slices where the corresponding planes don't contain a full slice: we constrain our slice selection to the portion that has a close-to-full slice correspondence in the reconstructed volume. If the plane correspondence is correct, the images will show the same anatomical features. Clearly our method catches the correspondence better than the reconstruction-first method. We can also glance at the registration performance of our registration method described in Section 3.3 from Figure 4.7.

## 4.3 Posterior Brain

External thermal and mechanical stimuli are detected and transmitted to the spinal cord by peripheral afferents with their cell bodies in the dorsal root ganglion. Within the spinal cord, the projection neurons further transmit the sensory information to various brainstem and thalamic nuclei. This ascending sensory information is also subjected to extensive processing by different mechanisms. The type of modulation that the study is interested in is from a network of descending pathways

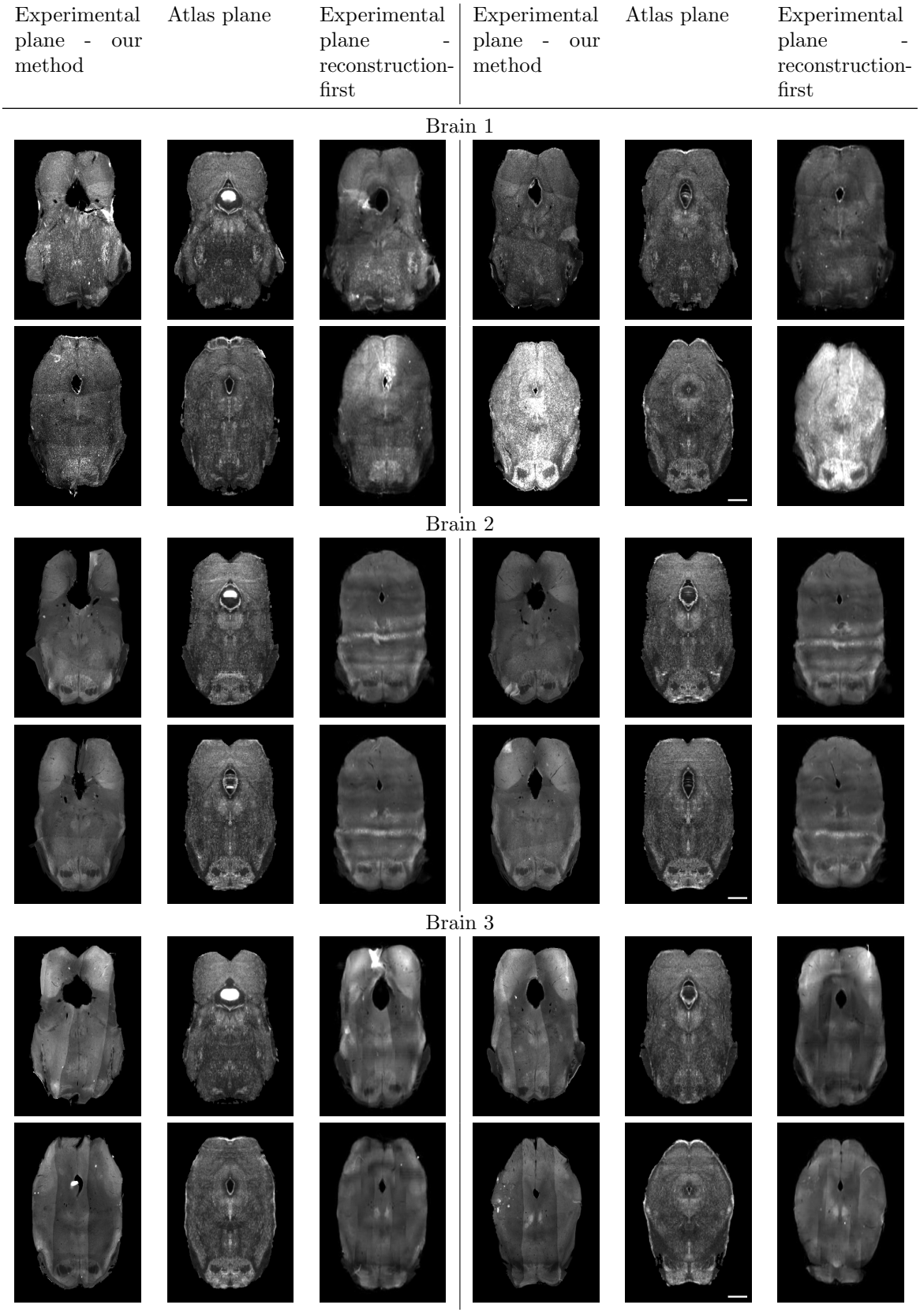



Figure 4.7: Dorsal raphe study results: experimental images and their corresponding atlas planes after registration. Intensity 2× in all images for visualization purposes. Scale, 1 mm.

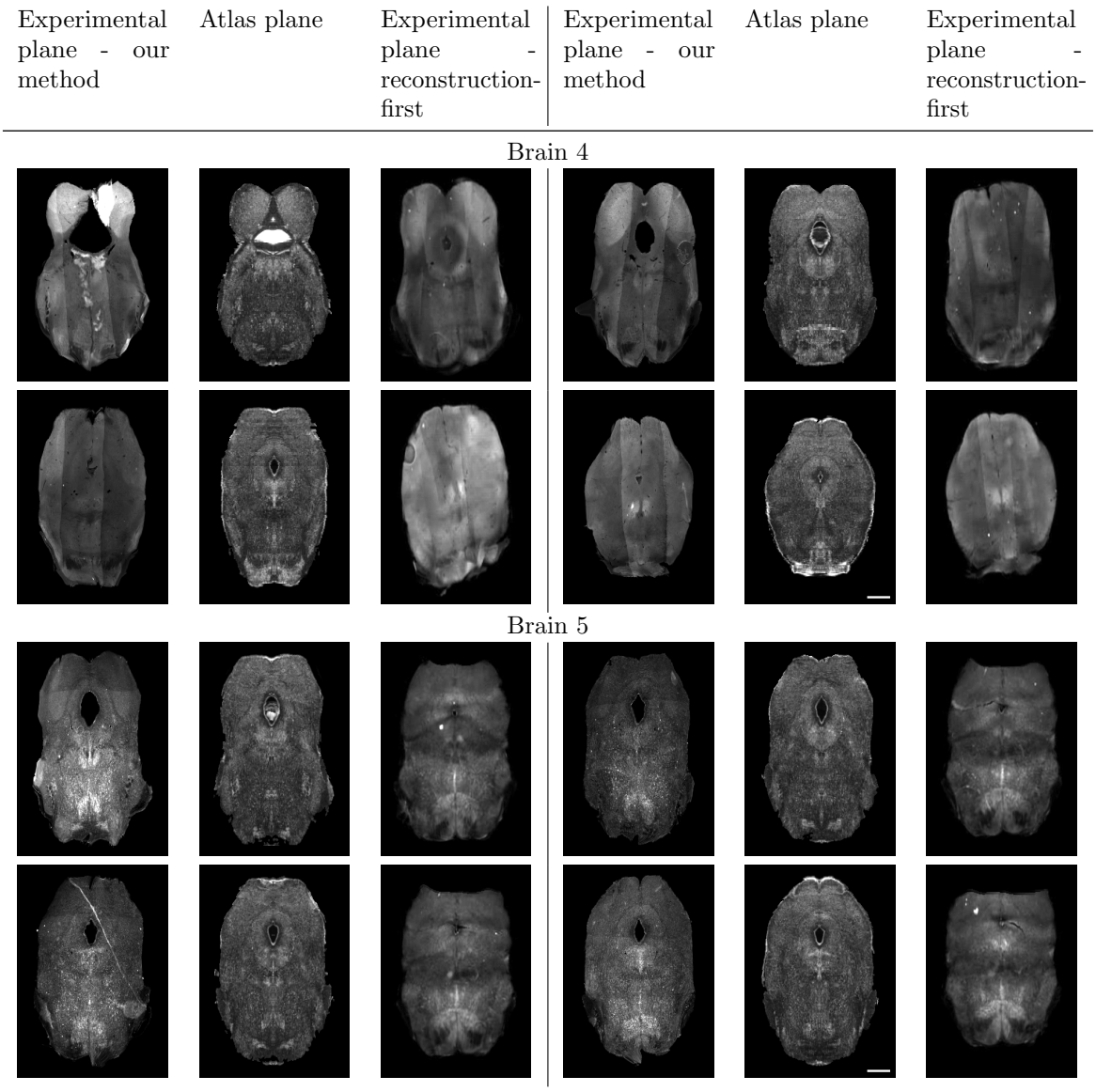



Figure 4.7: Dorsal raphe study experiments results: experimental images and their corresponding atlas planes after registration. Intensity 3× in experimental images, 2× in atlas images for visualization purposes. Scale, 1 mm.

projecting from cerebral structures to the spinal cord dorsal horn. This top-down pathway includes the anterior cingulate cortex, the hypothalamus and the amygdala that relay information to the midbrain periaqueductal grey, and further downstream to the rostral ventromedial medulla (RVM), including nucleus raphe magnus, nucleus reticularis gigantocellularis, and adjacent areas, to exert bidirectional control on the spinal cord nociception. Within the RVM, there are different types of neurons that have different pain modulatory functions. In this study, our brain mapping program is used to map three different kinds of spinal cord projection neurons in the RVM to the common Atlas coordinate and analyze them.

### 4.3.1 Dataset

The experimental datasets are posterior sectional brains consisting around 30 slices, because RVM is mostly in the posterior sections. We tested both on DAPI-stained and Nissl-stained brains and found that our algorithm does not work on DAPI-stained data for 30 slice posterior sectional brains. We analyzed the intermediate results and found:

1. DAPI stains all cells, not just neurons. Therefore the features shown in a DAPI-stained experimental brain does not correspond well to the atlas.

2. The posterior section consists of very strong features — regions in the cerebellum — in the upper half, while the features in the medulla in the lower half are smaller in number and have less contrast. With DAPI staining, the features in the upper half are still strong, but the features in the lower half are hardly visible, making our angle detection very hard.

3. An experimental sectional brain consists of only around 30 slices in the posterior brain. The lower half contrast visible with Nissl staining are now mostly missing with DAPI stain. Even a neuroscientist cannot determine a sectioning angle.

Therefore, we switched to Nissl stain instead. Even though the upper half features are still much stronger than the lower half features, because for determining sectioning angles, we compare upper and lower half features to slices in the neighborhood separately, the strength of features does not pose a problem as long as the lower half features show up.

### 4.3.2 Preprocessing

The experimental dataset contains a region called the paraflocculus, which is a lobe-like structure, as shown in Figure 4.8. The structure extends out from the brainstem and is connected with the brainstem through a thin tissue. The structure in 3D is a protruding tissue outside of the brainstem. Since it is only connected with the brain stem through a thin tissue, the structure can move backward or forward when an experimental mouse brain is taken out of the skull. Therefore the position of this structure is not reliable in determining the cutting angles of the microtome. We removed the structure

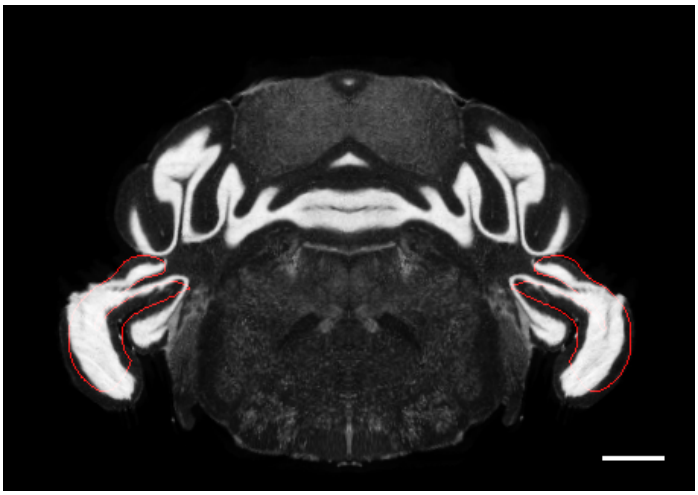


Figure 4.8: An atlas image containing the paraflocculus region, with the annotated paraflocculus region in the annotation volume circled in red.

consistently during preprocessing. Specifically, we removed the paraflocculus region and the fiber tracts region enclosed in the convex hull formed by the paraflocculus region, as well as the flocculus right near the paraflocculus which is also included in the "lobe". Because this lobe-like region is very distinguishable, an operator after some training would be able to complete the preprocessing task easily on the experimental images. Depending on the sectioning angle and the length of the experimental section, some more anterior slices contains cerebral cortex as well. Because cerebral cortex is not connected to the brain stem in the brain section of our experimental data, we removed the cerebral cortex as well. We removed the same regions in the atlas using the annotation volume. Since annotation is not very accurate like shown in Figure 4.8. The removal was sanity-checked by a neuroscientist. We chose a section in the atlas that well contains the volume that the experimental volumes would cover and removed the same regions as we did in our experimental volume. This atlas section was created once and was used as the reference for all the experimental brains in this study.

### 4.3.3 Registration

For the plane-wise mapping step (Section 3.2.1), we used the same implementation as the Dorsal Raphe dataset (Section 4.2). However, the best cutting angle returned by our program did not seem to be correct. Therefore we rotated the sectional atlas corresponding to our dataset and asked a neuroscientist to identify an angle that could be right. We checked the best matching slice indices for both the upper and lower half slice of every slice selected, and found the upper slices are matched correctly with our program, but the lower half slices were not. We further examined the intermediate steps and attributed the reasons to the following:

1. The posterior brain stem does not have any strong features. The few features are low-contrast compare to most features in other parts of the brain.

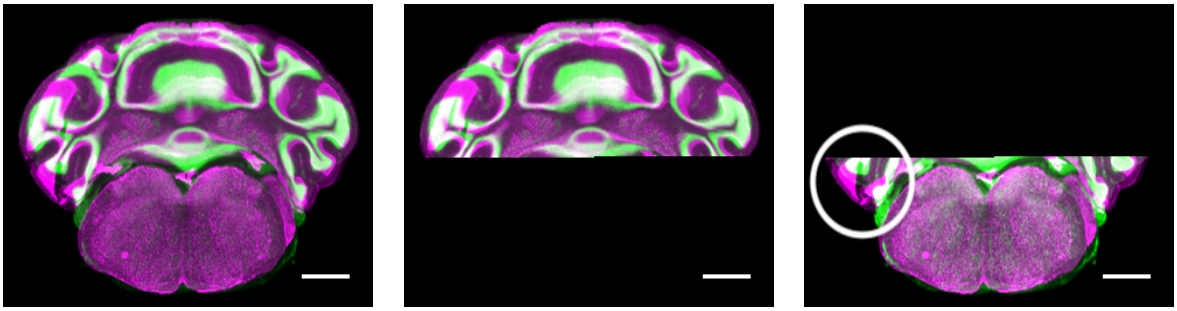
2. Using HOG difference to measure similarity does not require precise registration of two images and allows for distortion, but when all the features in an image is too small in magnitude, misalignment affects the result.

3. The lower half slices of these posterior brain slices are not aligned very well with our affine registration.

To illustrate the reasons above, we take an experimental slice in the posterior dataset as an example. To find the best cutting angle, a required step is to find the best matching half slices of an experimental slice for a given angle. Given an experimental image, we search a neighborhood of atlas slices and compute a HOG difference score between them. Figure 4.9 shows an overlay of this experimental image rescaled to the resolution of the atlas and a corresponding atlas slice affinely registered to it. The affine registration is performed as described in Section 3.1: we extract corresponding points on the contours of the two images, compute a similarity transform, and then apply additional horizontal scale and vertical scale that aligns the two shapes better. We then split the images in halves with the plane that cuts an unrotated atlas into the upper and lower halves and compute the HOG difference to their potential matches in the half atlases. An example of the upper half and lower half overlay of this experimental image to its corresponding atlas slice are also shown in Figure 4.9. We can see that after affine registration most corresponding features are close to each other. The algorithm correctly predicted the best matching upper half slice, but predicted another slice to be the best matching lower half slice.

When we examine slice matches, we see how many biological features correspond. In the lower half of the posterior images, we look mainly at the biological features in the brain stem region — the low intensity area in the bottom. The lower half slice in (C) showed corresponding biological features, but the cerebellum and its surrounding regions dominate the signal. With the circled area not well-aligned, the HOG difference of these two slices became large even if the two images had corresponding biological features. This shape difference may be because the cutting angle around the superior-interior axis is not zero. As a result, our experimental images were not symmetric. It may also because of inter-brain variations. Because this indentation difference is not large, and this misalignment is mainly near the outer contour, we used a simple approach — applying a shrinked lower half mask to the images. Thus the HOG difference does not include these misaligned regions caused by different tissue indentation. With this trick, the best matching half slices were found correctly.

We also tried to split these posterior images into the cerebellum part and the brain stem part instead of cutting the image in halves. However the brain stem is connected with the cerebellum through some fiber tracts tissue. Therefore in some slices, the boundary is not clear, and if we were



(A) An overlay of a posterior experimental image (magenta) and the corresponding atlas slice (green).

(B) An overlay of the upper half slices.

(C) An overlay of the lower half slices with not well-matched area circled in white.

Figure 4.9: Affine registration result of a posterior slice and half slices. Scale, 1 mm.

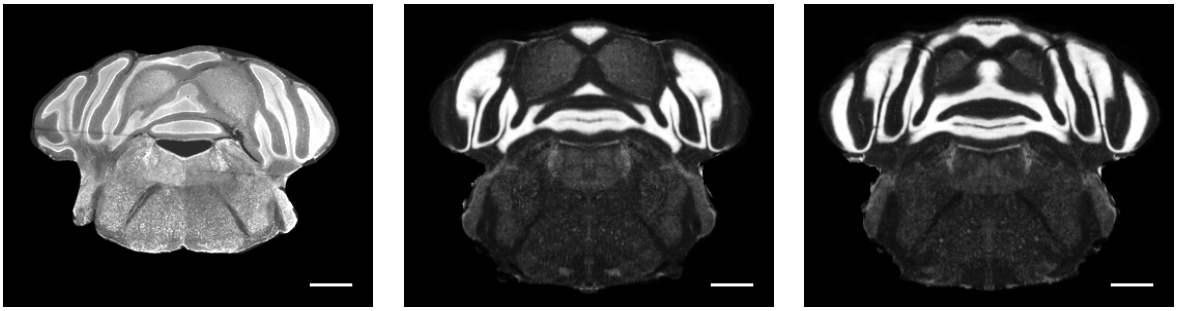
| Angle | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|
| | 0 | -5 | -4 | -3 | 1 | 5 |
| | -2 | -3 | -4 | 0 | 4 | 3 |
| | -4 | -4 | -4 | 0 | 6 | 4 |
| | -4 | 0 | -3 | 0 | 7 | 5 |
| | -1 | 1 | -1 | 1 | 10 | 5 |
| | -1 | 4 | 2 | 5 | 12 | 9 |
| | 3 | 8 | 6 | 6 | 16 | 13 |
| | 7 | 11 | 10 | 10 | 17 | 16 |
| | 11 | 14 | 13 | 14 | 20 | 18 |
| | 12 | 16 | 14 | 18 | 21 | 15 |

Table 4.2: Example upper and lower best matching half slice index difference of selected experimental image with different cutting angles.

to fix the misalignment this way, a manual fix is needed.

With this fix in place, we tried to find the best matching angle, but the angle was still not correct. Therefore we checked the best matching index difference between the upper and lower slices for every selected experimental image and every given angle. The numbers are shown in Table 4.2. Each column represents the index differences of best matching half slices under a given cutting angle. If we look at the numbers column-wise and do not include the last rows (more anterior slices in the posterior section), the numbers follow an expected pattern — from smaller angles to larger angles index differences are first mostly negative, then is around zero, and finally are positive. In this case, angle 7 gives an average index difference closest to zero and is indeed better than the other cutting angles when we examined the experimental brain manually. This means without the last four slices, our algorithm could detect the cutting angle correctly.

We further investigated the last slices and found out the cause of large index differences for the



Experimental slice | Best matching upper half slice | Best matching lower half slice

Figure 4.10: One of the last four experimental slices and its best matching upper and lower slice. Experimental slice resized to the resolution of the atlas. All images contrast adjusted for visualization. Scale, 1 mm.

last slices was the cerebellum. Figure 4.10 shows the best matching upper slice and the best matching lower slice of one of the last four slices given the best cutting angle was 7. From the figure, we can see the darker region (brainstem) is matched well in both images, but not the bright regions (cerebellum). The middle figure which is supposed to be the best matching upper half slice, the bottom center of the cerebellum is matched well but not the two sides. The right figure which is supposed to be the best matching lower half slice, the bottom center of the cerebellum does not match with the experimental image, but the two sides of the cerebellum are matched well. We guessed the anterior-posterior axis of the brain was bent, because while the brain stem seemed to match well with the experimental image, the features co-existed in the cerebellum of the experimental image appeared in two different planes in the same rotated atlas. We examined the mouse brain and found out the cerebellum was indeed loosely attached to the brainstem and was prone to displacement when the brain was taken out of the skull.

Figure 4.11 shows the annotated sagittal view of the mouse brain. The posterior slices are sectioned from the right and end around where the cerebellum (yellow region) ends. In the more posterior slices, brain stem is only in the lower half, as shown in Figure 3.1. In the most anterior slices in the posterior section, as shown in Figure 4.10, we can see the brainstem in both the upper and lower halves. Since our result shows the index differences are around zero for most slices from the posterior, and only the upper half is not matched well in the few anterior slices. Since cutting angle is relatively stable, our guess is that the mouse brain may bend in this brain near the connection between the cerebellum and the brainstem. One can see from the sagittal view that the cerebllum is connected with the brainstem (pink and orange in the posterior sections) through three pieces of tissues in this sagittal view. Gray regions are the ventricular systems which are empty space. In coronal views, the cerebellum is connected very loosely to the brainstem as shown in Figure 3.1 in the very posterior sections, and connected more to the brainstem through thicker tissues on the

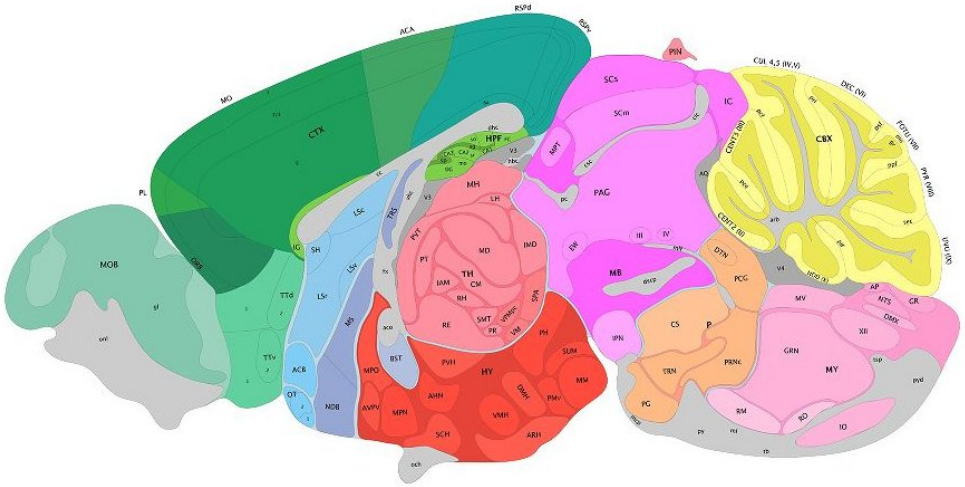


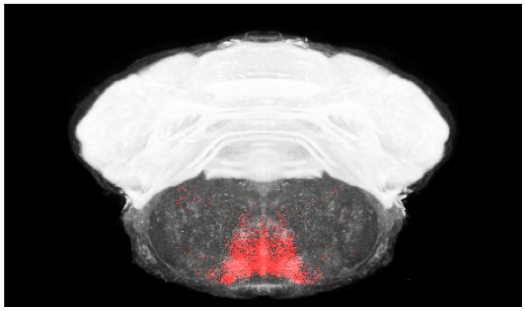

Figure 4.11: Annotated sagittal view of the ABA mouse brain. Left: anterior, right: posterior.

two sides of the fourth ventricle — the hollow structure in the middle — as shown in Figure 4.10. Because mouse brain is long, and the brainstem's size in coronal views suddenly drops around this region, 3D deformation is possible when the brain is taken out of the skull or during the handling process. In addition, there are large hollow structures in between the brainstem and the cerebellum, depending on the amount of perfusion and dehydration, the size of the ventricular systems may vary between the experimental brains and the atlas which may cause additional 3D deformation in the posterior section.

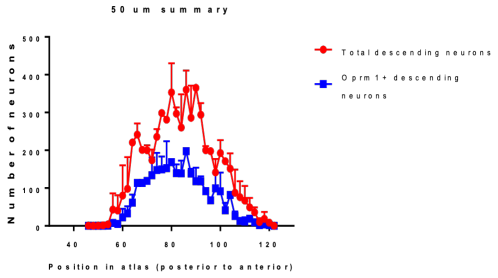
The neurons we are interested in in this study are mostly in the more posterior slices and in the brainstem. Since the brainstem are matched very well in all slices, we only made slight modification to only use the first two thirds of the slices to determine the cutting angle, but did not make further improvement.

#### 4.3.3.1 Pixel-wise Mapping

The cerebellum has much stronger contrast then the brainstem in these posterior slices. However, they are of equal importance. To make the salient point extraction more uniform, we separately treated the cerebellum and the brainstem using the atlas annotation. We extracted salient points in the two regions. Points in each region satisfy the spacing constraint. When two control points from different regions were too close, we kept the point from the brainstem, because in this study we need more precise registration in the brainstem.



(A) Average posterior brain atlas with all mapped neurons (red) plotted on top




(B) Preliminary analysis of posterior to anterior neuron distribution of the total descending neurons and the neurons from a specific pathway in each coronal atlas slice

Figure 4.12: Preliminary result of the posterior brain study.

### 4.3.4 Result

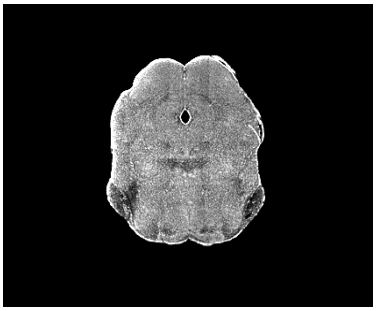
After we mapped all neurons in an experimental brain to the atlas, we generated an average coronal atlas image of the corresponding section and plotted the mapped neurons on the average atlas slice as shown by the red dots in Figure 4.12. The right chart shows a preliminary analysis of the anterior to posterior distribution of the total descending neurons (red) and neurons from a specific pathway (blue).

## 4.4 DAPI-stained Near-Full-Brain Dataset

While in the prior section our approach was not able to map a posterior section of a DAPI stained brain, when the brain dataset covers a longer section, our algorithm still works. It was used to analyze the neuron distribution in the hypothalamus and the brainstem at a whole brain level. The neuroscientists were interested in quantifying signals in small nuclei, therefore a precise 3D mapping is necessary.

### 4.4.1 Data

The experimental data was near full brain data with 50 µm sections where the more anterior slices, such as the olfactory bulbs, are not imaged, but about two thirds of the coronal slices are imaged. The dataset was stained with DAPI, which matches the Nissl stained atlas less well. However, unlike the brain section, this dataset had has more slices than in the posterior study, and many DAPI-stained images which did show some features, which enable the algorithm to function well.



(A) An example slice in the DAPI-stained brain (cortex missing).

(B) Best corresponding slice in the atlas returned by our algorithm.

(C) A Nissl-stained experimental slice in the similar region.

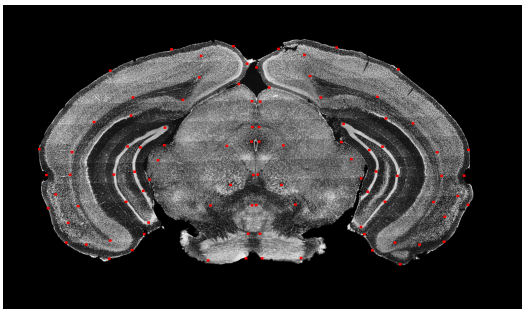
Figure 4.13: Comparison of features shown in a DAPI-stained experimental slice, an atlas slice, and a Nissl-stained experimental slice in a similar brain region.

### 4.4.2 Result

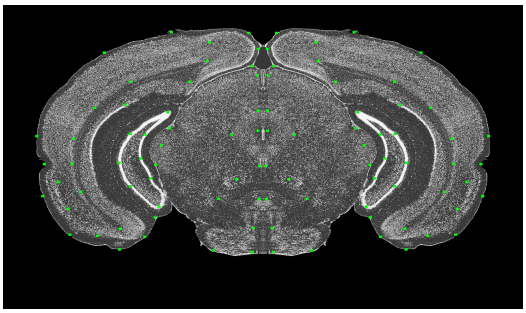
Because our registration algorithm relies on the geometric information to determine the initial rough searching space for each selected slice, we asked the neuroscientist to identify index numbers in the whole atlas which bounded the regions that were represented in the experimental brains. That step was needed because the olfactory bulbs were missing. The neuroscientist also made sure no image was flipped horizontally. Resolution of the experimental images was hard to extrapolate because of some arbitrary downsampling. Therefore we compared an experimental image to an experimental image that contain similar slice with known resolution and used the estimated resolution in our program. The cutting angles were returned successfully, and correspondences were satisfactory. We found that the more anterior slices are matched better than the very posterior slices. The cause may be the same as described in Section 4.3.1. Figure 4.13 shows an example slice of the matched DAPI-stained experimental brain and the corresponding atlas slice. The slice matches were verified by a neuroscientist. Visually the there were less features shown in the DAPI-stained slice in the brainstem. Even if a feature was stained, it is less clear than the Nissl-stained slices. This can be seen in the figure by looking for a sideways '8' in the bottom of the brainstem. This feature is much clearer in the Nissl stained atlas and experimental image.

## 4.5 High Resolution 2D Non-rigid Registration

While many anatomical studies analyze the mouse brain at a whole brain level, some studies are only interested in in-plane analysis. In this study [29], for example, neuroscientists registered 3 slices per animal to their corresponding 2D reference atlas image. Because of the large gap between slices, 3D analysis is not useful. In this study instead of the 3D isotropic 2011 or 2015 Nissl atlas that are



(A) An example high resolution experimental slice with manually selected points in red.



(B) Corresponding high resolution atlas slice in the Mouse coronal v1, 2008 atlas manually selected by neuroscientist and corresponding points marked in green.

Figure 4.14: An example of high resolution experimental image and its corresponding reference atlas coronal slice.

used in the previous studies, the neuroscientists identified corresponding high-resolution coronal atlas sections (Mouse coronal v1, 2008) from the ABA. While no sectioning angle difference was taken into account, only slices that are close to the coronal atlas sections's sectioning angle are kept. Since our method consists of two separable steps — plane-wise mapping and the pixel-wise mapping, if the corresponding slice is identified manually, we could directly use the second step to perform the 2D registration.

From the same lab where the study in [29] was performed, we obtained 10 coronal slices whose positions span throughout different regions of the mouse brain and their corresponding neuroscientist identified coronal reference atlas sections. In this lab, custom software was used to register the manually selected correspondent slices. The custom software used a stochastic hill-climbing method which optimizes a weighted sum of the overall image cross-correlation and the mapping distortion [33]. To improve the qualitative result from the registration method used by the lab, neuroscientists marked on average 127 pair of corresponding points in each slice and its corresponding high-resolution atlas section. An example is shown in Figure 4.14. Different from the landmark points selected in Section 3.7, the landmarks points are on feature boundaries but not necessarily feature corners. Therefore the neuroscientist is confident the landmark points are on the curve but may not be confident about the specific position of a landmark point on the curve. We ran our nonrigid registration algorithm with salient control points on these 10 pair of images and compared the TRE of using the custom software from the lab. We calculated the average TRE from each slice, and the boxplot of the average TRE per slice from these two methods is shown in Figure 4.15. Our method has a higher average TRE than the custom method used in the lab.

We further examined each pair of slices, the original points marked on the experimental images, the corresponding points marked on the manually selected reference images, the corresponding

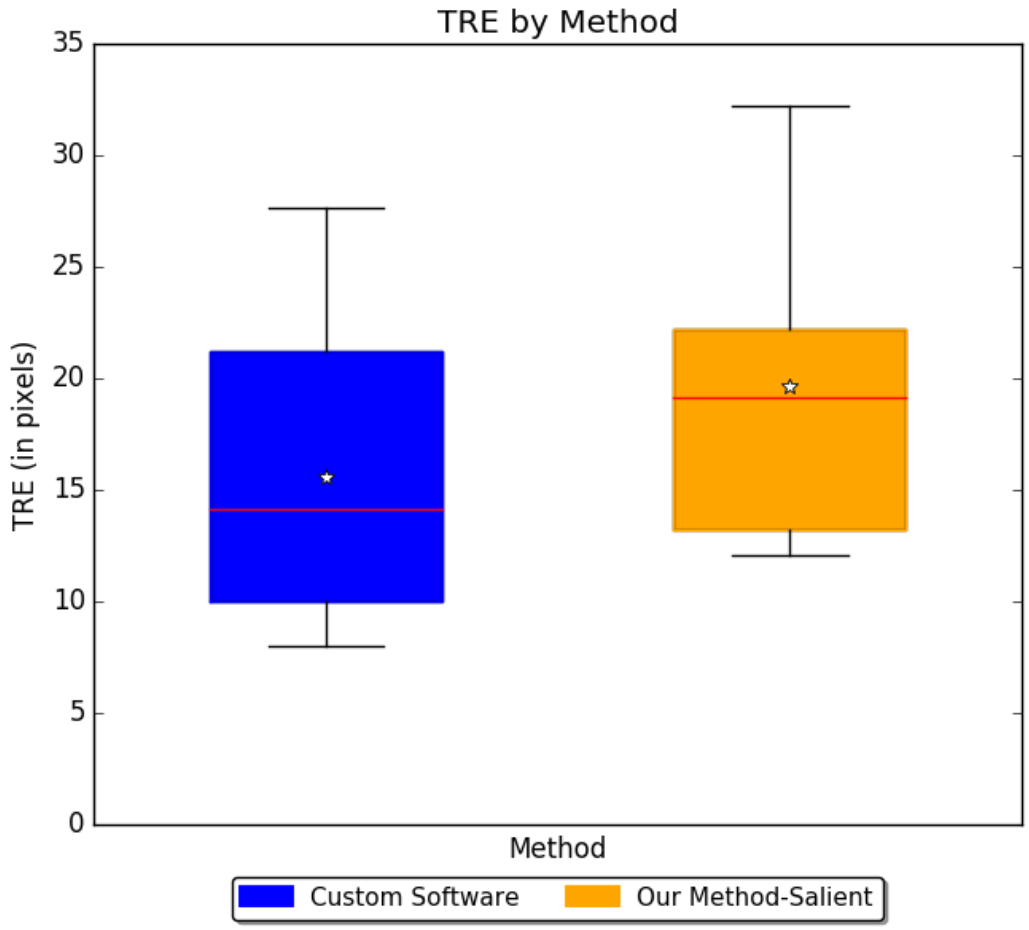
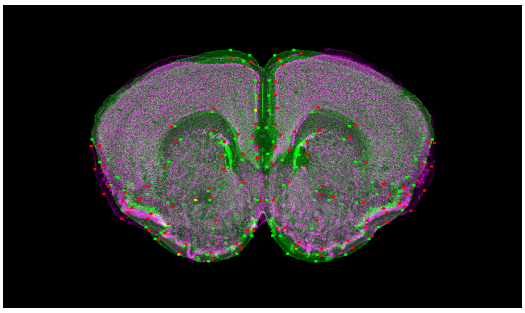


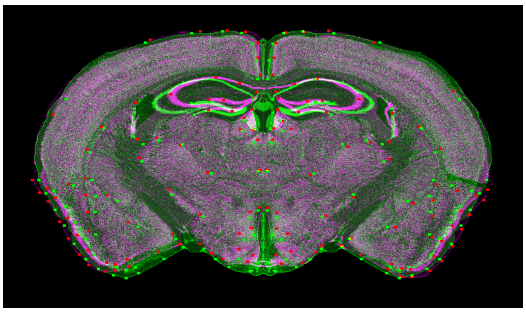

Figure 4.15: Boxplot comparing the TRE on 2D high resolution slices.

position of the original points on the warped experimental images with both methods, as well as the quality of the computed transformations. We found that for most of the slice pairs, there was not much difference between the results in terms of visual quality. But if there was some difference, our method performed better in registering tissue boundaries and when two corresponding features are not close after affinely registering the experimental image to the reference slice. Example pairs are shown in Figure 4.16. The image pair on the left column shows an example in which our method registers the image contours better. After affine registration, the contours of the two images are not aligned because affine transformation does not correct nonrigid deformation. In our result, the contours of the images are aligned better than that from the custom software. Internal structure alignment from the two methods are comparable. In the image pair in the right column, we can see after we bring the experimental and the reference image to the same coordinate with the affine transformation, the hippocampus in the two images are not well aligned. A nonrigid deformation is needed. The intensity-based custom software improves the alignment generally, but could not align the hippocampus, because the edges are thin and does not overlap much, therefore it is hard to achieve the optimial solution with a local iterative search. However our method is more feature-driven. We first place control points on the boundary of salient features. We then describe surrounding areas of these salient points with HOG descriptions, after which we search all the possible constrained displacement of every salient control points, minimize the energy function so that the HOG difference are minimized while smoothness is encouraged. Therefore with our method, the hippocampus is aligned much better. This example shows that our method generally aligns thin, non-overlapping or little-overlapping biological features better than an intensity-based method.

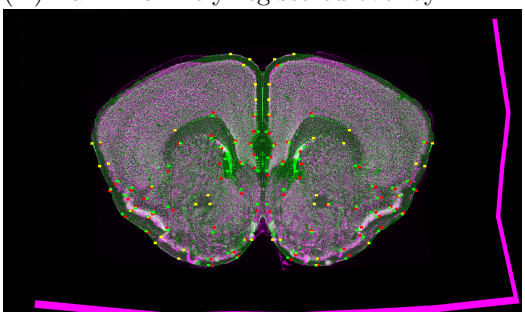
However, even though our method shows comparable qualitative result as and sometimes even better than the custom software, the boxplots in Figure 4.15 gives contradicting conclusion. We further examined the points placed by the neuroscientist. Most of them were placed on edges of biological regions that show contrast to surrounding regions. Some were placed inside blobs. Others were placed on the corners. Most of the edge points were placed based on relative distance. However, our method seems to give more TRE systematically than the custom software regardless of how we tuned the coefficients of our energy function to encourage smoothness. We manually checked the points which our method did much worse than the custom software and found that most of them were edge points. After confirming with the neuroscientist, we were informed that the corresponding points were selected using the warped image from the custom method as reference, and the goal was to add these points to improve the registration quality of the custom software. This is why the TRE from our method was not as small as the TRE of the custom software. An example is shown in Figure 4.17. Green and red points aligned much better in the custom software generated results, with many yellow points especially in the lower left part of the image. The example image was chosen because it shows a clear example of points that are biased. The experimental image is clearly rotated. However the points in the white circle are straight-up. The custom software did not correct



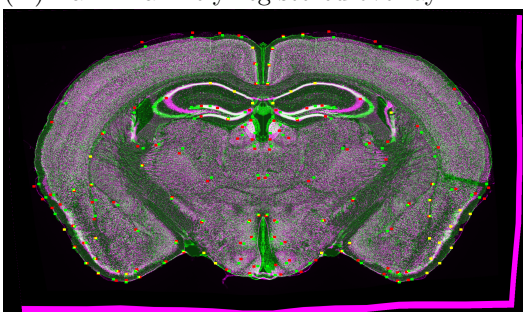
(A) Pair 1: affinely registered overlay.



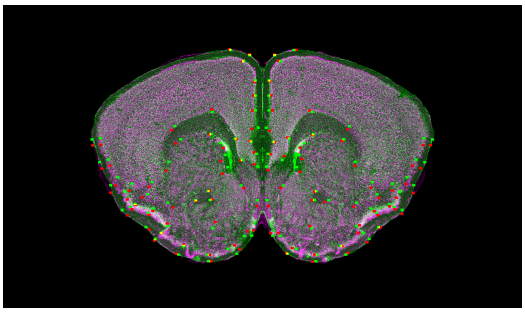
(D) Pair 2: affinely registered overlay.



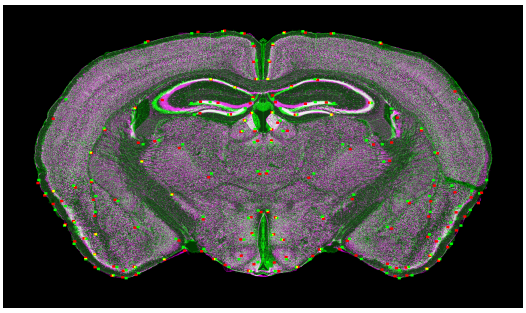
(B) Pair 1: overlay after warping with custom software.



(E) Pair 2: overlay after warping with custom software.



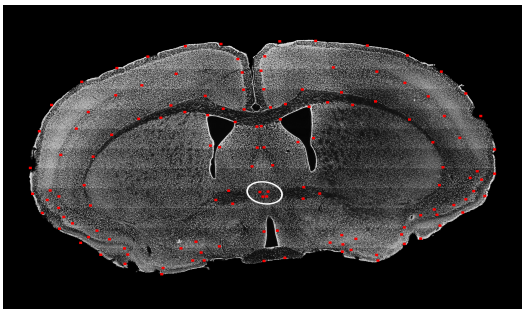
(C) Pair 1: overlay after warping with our method.



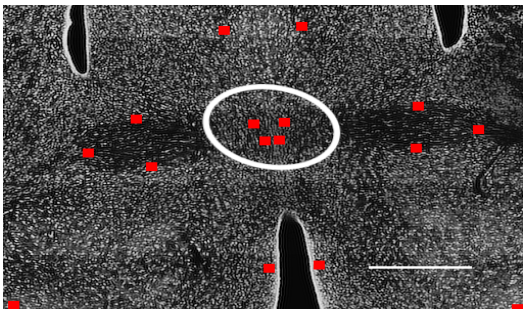
(F) Pair 2: overlay after warping with our method.

Figure 4.16: Examples of 2D nonrigid registration results. Warped experimental image (magenta) and reference image (green) overlay with different methods. Red points are the warped marked points, and the green points are the neuroscientist selected corresponding points on the reference image.

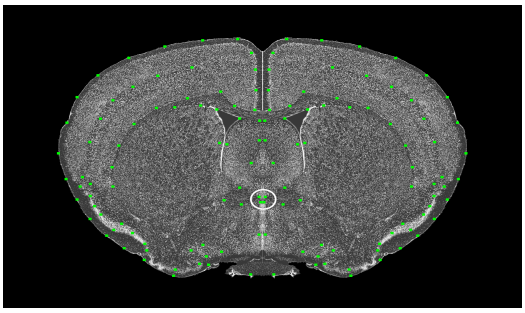
the rotation, but showed a better alignment in the circled region than our method, even though our method registered the two images better as the readers could tell from the zoomed-in view showing in Figure (H) where the feature in the bottom was aligned while the same feature in Figure (F) was still tilted. From the example, we can tell the points are biased, and therefore the results given in the Figure 4.15 are not a fair comparison between the two approaches.



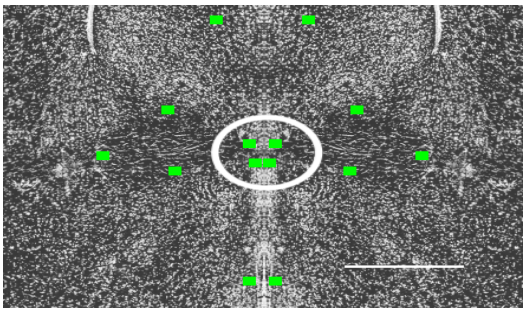
(A) Experimental image with marked points.



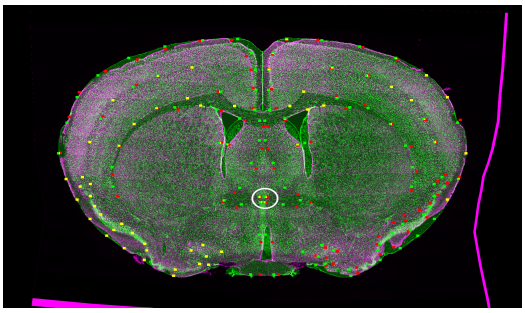
(B) Experimental image marked points zoomed-in view.



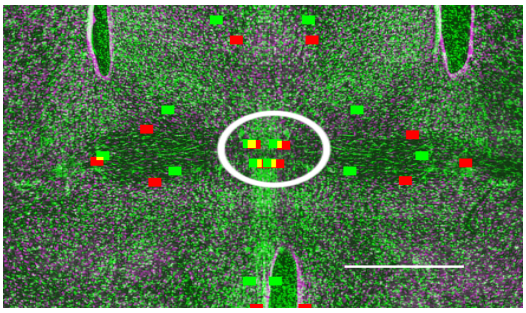
(C) Reference atlas image with marked points.



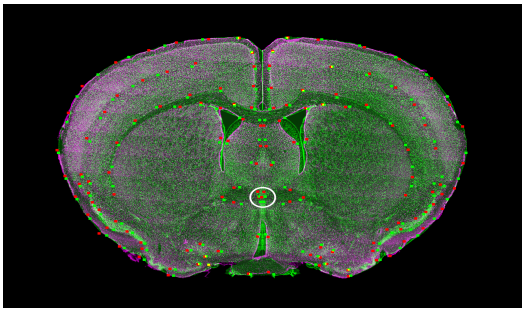
(D) Reference atlas image marked points zoomed-in view.



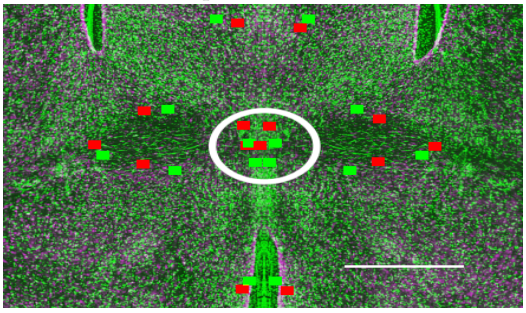
(E) Overlay after warping with the custom software.



(F) Overlay after warping with the custom software marked points zoomed-in view.



(G) Overlay after warping with our method.



(H) Overlay after warping with our method marked points zoomed-in view.

Figure 4.17: An example of false corresponding points. Warped experimental image (magenta) and reference image (green) overlay with different methods. Red points are the warped marked points, and the green points are the neuroscientist selected corresponding points on the reference image. If two points are aligned after transformation, the point will turn yellow (overlaying green and red). An example of false corresponding points is circled in white. Scale 1 mm.

# Chapter 5

# Discussion

## 5.1 Handling 3D Deformation

In Section 4.3.1, we found that there may potentially be a 3D deformation around the region where the brainstem and the cerebellum connects. In Section 4.4, we found that the more anterior slices are matched better than the most posterior slices. In Section 4.2 where brain sections were only about 1/6 to 1/7 of the full brain length, slices were matched very well. All the best matching slices were computed under the assumption that the cutting angle is consistent. Since all experimental mouse brains we have matched were adult mouse brains, there were not much structural variation between them. The differences we saw mostly came from the difference in staining — atlas shows more structure than the experimental brains, and ventricular system appearance — due to variation in the perfusion and dehydration processes. We have handled those differences with cycle consistency and the MRF formulation with tissue coherency. The one inconsistency we have not solved is the potential cutting angle change or 3D deformation we saw in the above two datasets.

In Table 4.2, we showed a matrix filled with upper and lower slice index difference given different cutting angles. When cutting angle equaled 7, the index differences suddenly increased to around 10, while the previous index differences were all close to zero. Table 5.1 shows the best matching upper and lower slice index difference from a full brain data at the angle which gives the minimum index difference. We used 35 experimental slices out of 220 slices, around every sixth image, to compute the best cutting angle. Even though the average index difference is close to zero, we can still see variation of the differences. The brain slices are ordered from posterior to anterior. The posterior index differences are mostly negative. In the middle, the differences are close to zero. Then the differences become larger towards the anterior slices. Because we assume a consistent cutting angle, we select this angle as the best cutting angle, and final plane-wise mapping gives good matches. However this data indicated there is more information we can collect. Our method has potential to generate more precise plane mapping and is able to handle 3D deformation or cutting angle variation.

| |
|---|
| 0 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -2 |
| -3 |
| -3 |
| -3 |
| -2 |
| 0 |
| 0 |
| -1 |
| 1 |
| 4 |
| 4 |
| 5 |
| 5 |
| 4 |
| 4 |
| 5 |
| 5 |
| 5 |
| 7 |
| 7 |
| 14 |
| 15 |

Table 5.1: Upper and lower best matching index differences for a full brain dataset at the best cutting angle computed by our algorithm.

## 5.2 New volume in the ABA

We have tested our algorithm both on the Nissl volume of ABA (2011) and ABA (2015). The Nissl volume of ABA (2015) is built from ABA (2011) by aligning the Nissl slices to the serial two-photon tomography (STP) brain. Since the Nissl volumes were put together from physically sectioned coronal histological slices, the volumes were not smooth, and after rotation, some edges were slighted jagged. Some scientists may be more interested in using the smoother STP volume. Therefore we did a small test on an experimental image which we computed the corresponding plane in the Nissl volume of ABA. Since the STP volume and the Nissl volume are aligned, we used the same plane in the STP volume as the corresponding STP image. We found:

1. The Nissl volume's background — non-tissue region — has small intensities around real tissue which causes problem on boundary alignment if not preprocessed. These small intensities are not consistent and sometimes similar to real tissue intensities. The STP brain atlas has nonzero intensity in all voxels, but these intensities are very low compared to those of real tissue and are consistent, therefore it is easier to preprocess the STP volume than the Nissl volume.

2. There are some regions in the STP brain that have very low-intensity values, similar to that of the ventricular systems, however those regions represent real tissues but not empty space.

3. The Nissl slice is slightly larger than the STP slice in the same plane.

4. The annotation atlas is accompanied with the Nissl volume, but not the STP volume.

5. The features in the Nissl slice also show up in the STP slice. The edges of features align. Therefore we are confident that our method works on both modalities, because HOG uses gradient features.

To correct for these effect and allow our code to be used we:

1. Used a very low intensity value to remove the background nonzero intensities in the STP brain slice to create a good contour. This contour is important since our energy function consists of a symmetric difference term that aligns the tissue contours. In the STP slice, a region that is not an empty region can have a low intensity similar to that of the aqueduct — an empty space. This issue may cause problem in preprocessing and needs special attention.

2. Removed the cortex region manually on the STP brain slice. The cortex of the Nissl atlas slices was removed first with annotation and then sanity-checked and thoroughly removed manually. This process was needed because the cortex of the experimental slice was mostly missing.

With the same parameter and code, we were able to register these two slices. Preliminary assessment with the neuroscientists we worked with was positive: they thought the quality of the registration was similar to registering to the Nissl volume, also pointed out places it could be improved (features which

were not aligned well). For this experiment both the Nissl and the STP volume needed preprocessing as mentioned above. Preprocessing the STP volume was easier because of its high signal to noise ratio. We did not test the plane-wise mapping part of our algorithm on the STP volume, nor did we thoroughly test the 2D non-rigid registration. But from this small experiment, we think with small modifications our algorithm can map experimental volumes to the STP volume of ABA.

## 5.3 A Deep Learning Attempt

Machine-based methods are becoming popular in the biomedical field. While many segmentation and detection tasks achieve high accuracy performance, learning-based registration results still lag behind the traditional registration methods in existing research. Nonrigid registration tasks are even harder because it requires a good formulation of the network including the input and output. The general bottleneck of using machine learning to do biomedical tasks is the lack of training data. It requires at least some transfer learning if we want to use a ready-trained network on natural images. Sometimes even transfer learning fails because biomedical images are very different from natural images. However, the advantage of a pretrained network is that the inference time is small — usually less than 1 second — comparing to the computation time to register a pair of images nonrigidly — several minutes in our case. Because in biomedical applications, there are usually hundreds of images to map, having a fast tool is appealing.

### 5.3.1 The ground truth

The first problem we faced was the lack of ground truth. In registration tasks, the desired output is how to warp the moving image so that it is aligned with the fixed image, the warp field to be more specific. Unlike segmentation that a neuroscientist could quickly circle the target structure out, the problem with nonrigid registration is that it is time-consuming to even ask a neuroscientist to warp a single image to its reference. Accuracy over weighs the importance of speed in biomedical tasks. Our histological images were mostly distorted, and even if the reference image looks very similar, multiple control nodes needed to be dragged in order to align the two images perfectly. Because we had developed a non-learning-based method to nonrigidly register an experimental image to its reference image, we utilized the computed warp fields generated with the uniform grids as the ground truth even though they are not perfect.

### 5.3.2 Network

Because we had success with the dilated convolutional network on the segmentation task with limited data, we used a similar network built with dilated convolutional layers. The input was both the fixed image and the moving image. We merge them together as a 2-channel image. We set the

output as the deformation field. Specifically the output is also a 2-channel output. To link the output directly to our non-learning-based registration pipeline, we describe the deformation field output as a query field — what coordinate locations we should interpolate to fill in a pixel in the warped image. Because convolutional neural network is translation-invariant, we use the relative query field — relative offset between the coordinates of original pixels we interpolate the current pixel intensity from and the current pixel — instead of the absolute coordinate. Because the aim is to align corresponding structures regardless of its location, registration should be translationally-invariant. Therefore the receptive field of a neuron at any level of our network does not need to be on the same scale as the size of the input image. As a result, we no longer need as many layers as in the segmentation network we built for aqueduct segmentation. But in order to correct distortions of various sizes, the largest receptive field of the registration network still needs to be on the scale of largest potential distortion. Therefore we removed a couple of layers from the original network, and the new configuration is given in Table 5.2. The images were resized to $256 \times 256$ whose original size was around $300 \times 400$. The largest receptive field is $65 \times 65$ on the same scale as the spacing we used to implement our method with uniform grid. Since we used image patches that is $2 \times 2$ the spacing, with the resizing, the largest receptive field functioned similarly. Same as before, the activation function is leaky ReLU.

### 5.3.3 Data augmentation

Similar to the segmentation task in Section 4.2.4.1, we did heavy data augmentation by applying additional nonrigid deformation to the deformation field and the image pairs and generated around 5000 image pairs and their warp fields. Because the query field defines how an image is warped, the output needs to be more precise than the segmentation task which is a binary output and a postprocessing can be easily added.

Due to the limited number of experimental brains, we tried to create synthetic data using the atlas directly. Even though the experimental dataset and the atlas are both Nissl-stained, they are not monomodal because the imaging settings and staining strength are different. To compensate and generate more "plausible" data, we applied different intensity and contrast alternation to the atlas images, then warp them nonrigidly, and pair them with their original atlas images. We replicated this technique but used the experimental images as the starting point, rather than the atlas. We trained our network using different combinations of these three datasets, and found out the model trained on the original dataset generated with the real experimental and atlas pairs with the traditionally computed deformation field as the ground truth worked the best. It makes sense because neither of the synthetic datasets made from solely experimental images or atlas images mimic the different image characteristics between the experimental images and the atlas well.

| Layer | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Convolution | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $3 \times 3$ | $1 \times 1$ |
| Dilation | 1 | 2 | 4 | 8 | 16 | 32 | 1 |
| Truncation | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Receptive field | $3 \times 3$ | $5 \times 5$ | $9 \times 9$ | $17 \times 17$ | $33 \times 33$ | $65 \times 65$ | $65 \times 65$ |
| Output Channels | 32 | 32 | 32 | 32 | 32 | 32 | C |

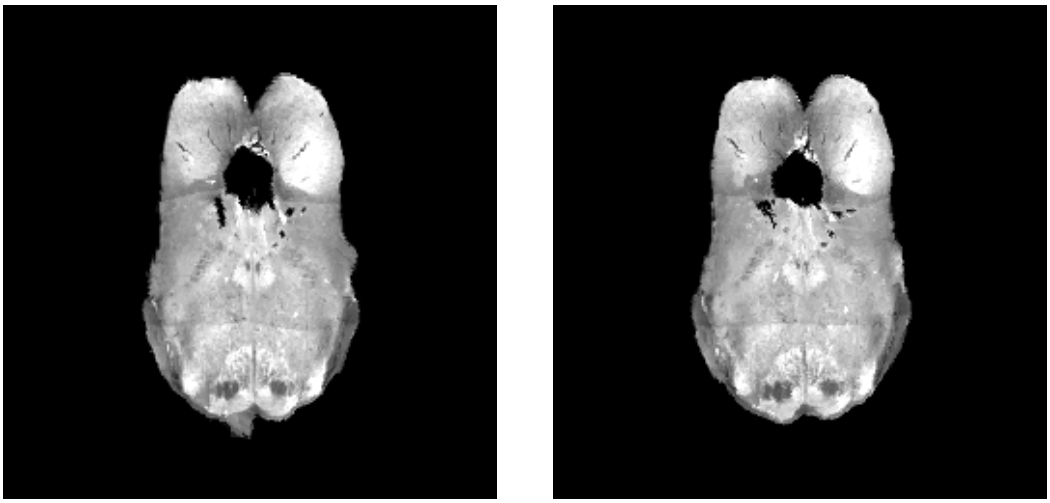
Table 5.2: Nonrigid registration network architecture. C is the number of channels in the input image.

### 5.3.4 Loss function

The loss function is formulated according to the characteristics of the histological image and the formulation of our network. Because the background pixels are all zeros. Mapping a zero-intensity pixel to a zero-intensity pixel should not be penalized even if the relative query coordinate is different from the "ground truth". However mapping a zero-intensity pixel to a nonzero pixel should be penalized and vice versa. Therefore instead of the commonly used average Euclidean distance, loss function is by the summation Euclidean distance between the computed relative query field and the "ground truth relative query field" over the union of the non-zero pixels in the warped image and the reference image divided by the intersection of the non-zero pixels in the warped image and the reference image. Therefore it is an "average" loss, and overlap of the warped image and the reference image is encouraged. We call it the "masked flow error". A slight regularization is also added so that the warp is not very large.

### 5.3.5 Result

The average loss in pixels was reduced to around 2.5 on our validation images after training. Since the input images are downsized to $256 \times 256$, and the original image were affinely registered to its corresponding altas section in the rotated atlas. Since an atlas image in the unrotated atlas is $320 \times 456$ pixels. 2.5 pixels is at most around 4.5 pixels on the atlas scale — 25 µm per pixel — much smaller to the expert error which was previously measured to be 9 pixels. We tested the trained network on the corresponding experimental and atlas image pairs computed by our algorithm. The experimental images are globally transformed to the atlas image coordinates because the network is for the nonrigid registration only. We applied the query field output to the original experimental images and obtained the final results. The transformed images are mostly aligned to the corresponding atlas qualitatively, and the results are visually satisfactory which an average difference of around 6 pixels (maximum around 11 pixels in the atlas scale) to the "ground truth" flow generated by our uniform grid method. It is similar to that of the expert error. However, after we examined the flow field and qualitatively compare to the flow field computed with MRF, we found the flow field output of the network is not as smooth. Figure 5.1 shows an example registration result of an image pair that we tested with the trained network compared with the traditional method generated



(A) "Ground truth" result from our uniform grid based MRF nonrigid registration.

(B) Result from the network which is qualitatively similar to the "ground truth".

Figure 5.1: An example registration result from the deep learning network — the "ground truth" warped image and registration result from applying the query field generated by the trained network.

result. The registration result is qualitatively similar to that generated with our uniform grid based MRF nonrigid registration and may look even better — contour of the warped image looks smoother. However, the flow field, as shown in Figure 5.2, is not smooth which is not desired.

### 5.3.6 Summary

Using deep learning to solve registration tasks is hard, especially for multi-modal registration problems. The first limitation is the lack of ground truth data. Unlike mono-modal problems where you can generate training data by warping images with known transformations, for multi-modal registration, there is no way a ground truth is available unless the multi-modal images were in the same coordinate to begin with. However, if this is the case, no registration would be needed. To address the problem, we tried different methods before using the traditional method to generate "ground truth" training data. We trained our network on simulated training data generated by warping atlas images with known transformations, simulated training data generated by warping experimental images with known transformations, and a mix of both. None of these training data generated good quality results. Therefore before a deep learning registration method can be effective, this problem with creating training data for multi-modal images needs to be addressed.

The second limitation is the lack of data. We only had five experimental sectional brain total with around 30 slices for each brain from the Dorsal Raphe study [53]. We allocated several images out from each brain and data augmented to 50 pairs as the validation data. We also allocated several

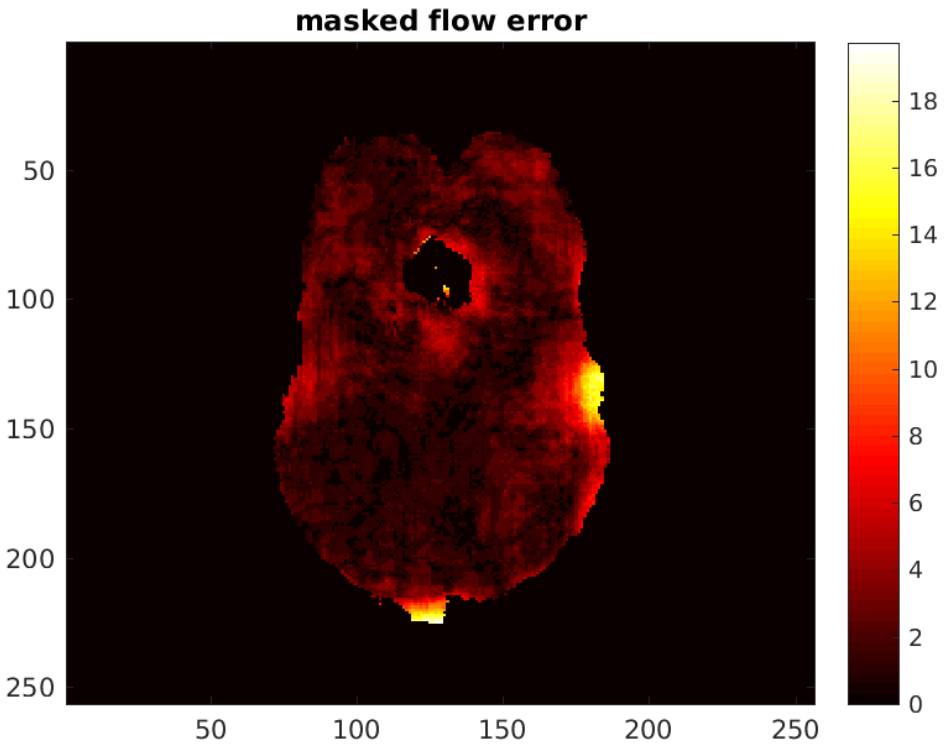



Figure 5.2: Masked flow difference between the "ground truth" query field and the network generated query field.

images out from each brain as the test data. We data augmented all other slice pairs to 5000 images as the training data. Our dataset size is small to begin with. Before we used our dataset to train, we searched available registration dataset online, but did not find any public histological mouse brain registration dataset. When we tested the trained network on slices in the same section in a full brain dataset, whose slices the network had not previously trained on, the result was worse than that of a brain which contributed to our training data. It means if the training data has not cover large variability, it perhaps is not scalable.

When analyzing the unsmooth results and brainstorming about solutions, we realized that our network was reverse to the traditional method. In a traditional nonrigid image registration method, we first fix large distortions, and then correct smaller distortions iteratively. However in our network and many existing convolutional neural networks, the receptive field of a filter is small in the first layer. In the following layers, the receptive fields would grow larger and larger, which is reverse to the process of the traditional approach. In a convolutional neural network, the smoothness of the result is dependent on the number and quality of the training data and the quality of training, one may want to design the network better to have the smoothness "hard-coded" to reduce this dependency. The unsmoothness may also be a result from not having enough training data which is hard to resolve for nonrigid biomedical image registration tasks.

# Chapter 6

# Conclusion

Histological sectioning is the most commonly used method to investigate organizations of normal and diseased brains. Individual brain variations and distortions and intensity inconsistencies caused by sample preparations make aligning histological brain slices to a reference a challenging task for both experts and computer algorithms. To address these challenges, we put together a direct approach to solving the mapping problem between a 2D histological sequence and a reference volume that allows us to determine the best corresponding slice for each experimental slice before attempting any nonrigid alignment. It uses the squared difference of HOG L2 norm as the image comparison metric and the average matching index difference between half-images to create a sectioning angle measurement. The HOG metric enables image similarity comparison without the need of the deformable registration. This produces a robust framework that leverages brain structural characteristics and symmetry to determine the cutting angle and matching slices without initial reconstruction.

Avoiding reconstruction improves accuracy by preventing z-shift or banana problems as validated by our comparison experiments. Interestingly, the results from the comparison experiment between the reconstruction-first method and our method show that using sectional reconstruction for registration still introduces small errors. These methods must compromise between thinner sections, with less z-shift issues, and thicker sections that contain better matching information. As a result, our method has better accuracy for registrations of sections with only 1/7 of the full brain. Our method works for datasets that include severe artifacts as well, because when finding the plane-wise mapping, we could avoid choosing images with severe artifacts and interpolate from best-matching planes of good quality slices to get the best-matching planes of those slices.

In 2D nonrigid registration, we first augmented the standard MRF on medical image registration to model accumulated tension when deforming tissues to more naturally deal with easily-deformed cavities throughout the brain. This requires us to use squared distance pairwise term and pass simulated stress across iterations. We further improved our 2D nonrigid registration algorithms by placing control points on the boundary of significant features and check if the point is consistent in

both images through cycle consistency check. The cycle consistency check makes our 2D nonrigid registration robust to artifacts. For slices with a large missing tissue we could preprocess the reference plane to remove the corresponding tissue manually before registration. With our method a neuroscientist can fully analyze all their experimental images regardless of their quality.

Since our method is mostly automatic, and the accuracy is similar to or better than an expert neuroscientist even for datasets where many slices are corrupted, we have successfully used our method to map multiple brain datasets to the ABA, making multi-brain data analysis possible and accurate. We are also collaborating with different labs on registering their data with different applications. Our experience working on different biological datasets and collaborating with different labs have shown that computer vision and image processing technologies can be widely used to automate many processes that biologists or neuroscientists routinely do. Our success came from starting our collaboration early, making it possible to find the problems caused by the unique image generation/capture used in brain slice imaging, and developing tools to deal with them.

Our pipeline is flexible which makes it easy to make changes and add new methods to deal with data-specific issues. For example, in Section 4.2.4.2, we add an extra aqueduct segmentation in our pipeline and add this information in the MRF energy function to make registration around the aqueduct region more accurate. In Section 4.3, we added a dilation mask to improve results on the posterior brain datasets. Our work can be easily extended to cover registration tasks in many other brain regions.

While our current processing pipeline is effective, further research can extend both its functionality, and its performance. For example, as indicated by Section 5.1, the plane-wise mapping seems to have sufficient signal to detect change in slicing angle caused by deformation of the brain during slicing. It would be interesting to extend the tool to capture this effect to make it more accurate. Improving the accuracy of the nonrigid registration also seems like a promising area.

Using the tool and tuning its parameters to optimize results would also be easier with improvements in the basic algorithms which power this pipeline. One promising area is in 2D nonrigid registration. We put together a framework that computes salient control points with cycle consistency to reduce some unsmoothness artifacts we saw in the uniform-grid generated results. But the method requires computing and matching the HOG profiles of a candidate salient point with every point in the neighborhood. While adding a HOG cache greatly improved this algorithm, further improvement in this area is still possible. We hope this work, and its future improvements will continue to aid our quest to better understand brain function.

# Appendices

# Appendix A

# Allen Brain Atlas Resources

While the Allen Institute provides a user-friendly API, running our code or other program to map an experimental volume or image to the atlas requires downloading the atlas volume or image. The Nissl volume we use to map sequential histological images are downloadable through link: help.brain-map.org/display/mousebrain/API. Detailed instructions are provided. The annotation volume is also available through the same link. The Nissl volume stores the pixel values, and the annotation volume stores the ID number of the region the pixel belongs to. A region structure graph is provided on the same webpage under "Downloading an Ontology"s Structure Graph". A JSON or a hierarchical XML document is provided. We use the JSON file (http://api.brain-map.org/api/v2/structure_graph_download/1.json). Anatomical structures are ordered in a hierarchical manner, we parsed the JSON file with python (code available on our project link) and read the generated text file in MATLAB and saved it in a usable manner that every subregion enclosed by a region can be directly returned through the region's ID number. (.mat file and code available on our project link). You are welcome to use any of those through citing our paper [66].

# Bibliography


[1] (2011). In situ hybridization data production. http://help.brain-map.org/download/attachments/2818169/ABADataProductionProcesses.pdf.

[2] (2015). Allen institute for brain science. brain atlas api. help.brain-map.org/display/mousebrain/API.

[3] (2018). Organization and function of neural circuits in the mouse. http://web.stanford.edu/group/luolab/Research/Section3.html.

[4] Abdelmoula, W. M., Carreira, R. J., Shyti, R., Balluff, B., van Zeijl, R. J., Tolner, E. A., Lelieveldt, B. F., van den Maagdenberg, A. M., McDonnell, L. A., and Dijkstra, J. (2014a). Automatic registration of mass spectrometry imaging data sets to the brain atlas. *Analytical chemistry*, 86(8):3947–3954.

[5] Abdelmoula, W. M., Škrášková, K., Balluff, B., Carreira, R. J., Tolner, E. A., Lelieveldt, B. P., van der Maaten, L., Morreau, H., van den Maagdenberg, A. M., Heeren, R. M., et al. (2014b). Automatic generic registration of mass spectrometry imaging data to histology using nonlinear stochastic embedding. *Analytical chemistry*, 86(18):9204–9211.

[6] Adler, D. H., Pluta, J., Kadivar, S., Craige, C., Gee, J. C., Avants, B. B., et al. (2014). Histology-derived volumetric annotation of the human hippocampal subfields in postmortem mri. *Neuroimage*, 84:505–523.

[7] Bagci, U. and Bai, L. (2010). Automatic best reference slice selection for smooth volume reconstruction of a mouse brain from histological images. *IEEE Transactions on Medical imaging*, 29(9):1688–1696.

[8] Bakker, R., Tiesinga, P., and Kötter, R. (2015). The scalable brain atlas: instant web-based access to public brain atlases and related content. *Neuroinformatics*, 13(3):353–366.

[9] Belongie, S., Malik, J., and Puzicha, J. (2000). Shape context: A new descriptor for shape matching and object recognition. In *Nips*, volume 2.

[10] Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585.

[11] Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, 1(3):185–216.

[12] Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., and Shen, D. (2018). Deep learning based inter-modality image registration supervised by intra-modality similarity. In *International Workshop on Machine Learning in Medical Imaging*, pages 55–63. Springer.

[13] Chen, Q. and Koltun, V. (2015). Robust nonrigid registration by convex optimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2039–2047.

[14] Chen, Q., Xu, J., and Koltun, V. (2017). Fast image processing with fully-convolutional networks. In *IEEE International Conference on Computer Vision*.

[15] Chew, L. P. (1989). Constrained delaunay triangulations. *Algorithmica*, 4(1-4):97–108.

[16] Cifor, A., Bai, L., and Pitiot, A. (2011). Smoothness-guided 3-d reconstruction of 2-d histological images. *NeuroImage*, 56(1):197–211.

[17] Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S., and Jefferis, G. S. (2016). Nblast: Rapid, sensitive comparison of neuronal structure and construction of neuron family databases. *Neuron*, 91(2):293–311.

[18] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

[19] Dauguet, J., Delzescaux, T., Condé, F., Mangin, J.-F., Ayache, N., Hantraye, P., et al. (2007). Three-dimensional reconstruction of stained histological slices and 3d non-linear registration with in-vivo mri for whole baboon brain. *Journal of neuroscience methods*, 164(1):191–204.

[20] Dorocic, I. P., Fürth, D., Xuan, Y., Johansson, Y., Pozzi, L., Silberberg, G., et al. (2014). A whole-brain atlas of inputs to serotonergic neurons of the dorsal and median raphe nuclei. *Neuron*, 83(3):663–678.

[21] Geha, P. Y., Baliki, M. N., Harden, R. N., Bauer, W. R., Parrish, T. B., and Apkarian, A. V. (2008). The brain in chronic crps pain: abnormal gray-white matter interactions in emotional and autonomic regions. *Neuron*, 60(4):570–581.

[22] Gibson, E., Gómez, J. A., Moussa, M., Crukley, C., Bauman, G., Fenster, A., and Ward, A. D. (2012). 3d reconstruction of prostate histology based on quantified tissue cutting and deformation

parameters. In *Medical Imaging 2012: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 8317, page 83170N. International Society for Optics and Photonics.

[23] Glaser, E. M. and Van der Loos, H. (1981). Analysis of thick brain sections by obverse—reverse computer microscopy: Application of a new, high clarity golgi—nissl stain. *Journal of neuroscience methods*, 4(2):117–125.

[24] Glocker, B., Komodakis, N., Tziritas, G., Navab, N., and Paragios, N. (2008). Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, 12(6):731–741.

[25] Goubran, M., Crukley, C., de Ribaupierre, S., Peters, T. M., and Khan, A. R. (2013). Image registration of ex-vivo mri to sparsely sectioned histology of hippocampal and neocortical temporal lobe specimens. *Neuroimage*, 83:770–781.

[26] Haber, E. and Modersitzki, J. (2006). Intensity gradient based registration and fusion of multi-modal images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 726–733. Springer.

[27] Hammelrath, L., Škokić, S., Khmelinskii, A., Hess, A., van der Knaap, N., Staring, M., et al. (2016). Morphological maturation of the mouse brain: An in vivo mri and histology investigation. *NeuroImage*, 125:144–152.

[28] Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416):391–399.

[29] Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., Yamashita, S., Bienkowski, M. S., Zingg, B., Zhu, M., et al. (2016). The mouse cortico-striatal projectome. *Nature neuroscience*, 19(8):1100.

[30] Ibanez, L., Schroeder, W., Ng, L., and Cates, J. (2005). The itk software guide: updated for itk version 2.4.

[31] Jefferis, G. S., Potter, C. J., Chan, A. M., Marin, E. C., Rohlfing, T., Maurer, C. R., et al. (2007). Comprehensive maps of drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell*, 128(6):1187–1203.

[32] Ju, T., Warren, J., Carson, J., Bello, M., Kakadiaris, I., Chiu, W., et al. (2006). 3d volume reconstruction of a mouse brain from histological sections using warp filtering. *Journal of Neuroscience Methods*, 156(1):84–100.

[33] Juels, A. and Wattenberg, M. (1996). Stochastic hillclimbing as a baseline method for evaluating genetic algorithms. In *Advances in Neural Information Processing Systems*, pages 430–436.

[34] Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. (2010). Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205.

[35] Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1568–1583.

[36] Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97.

[37] Lee, D.-T. and Schachter, B. J. (1980). Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242.

[38] Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–176.

[39] Li, H. and Fan, Y. (2017). Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv preprint arXiv:1709.00799*.

[40] Li, H. and Fan, Y. (2018). Non-rigid image registration using self-supervised fully convolutional networks without training data. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1075–1078. IEEE.

[41] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

[42] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

[43] Luo, L. (2015). *Principles of neurobiology*. Garland Science.

[44] Maintz, J. A. and Viergever, M. A. (1998). A survey of medical image registration. *Medical image analysis*, 2(1):1–36.

[45] Majka, P. and Wójcik, D. K. (2016). Possum—a framework for three-dimensional reconstruction of brain images from serial sections. *Neuroinformatics*, 14(3):265–278.

[46] Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10):761–767.

[47] Maurer, C. R., Qi, R., and Raghavan, V. (2003). A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270.

[48] Mertzanidou, T., Hipwell, J. H., Reis, S., Hawkes, D. J., Bejnordi, B. E., Dalmis, M., et al. (2017). 3d volume reconstruction from serial breast specimen radiographs for mapping between histology and 3d whole specimen imaging. *Medical Physics*.

[49] Miao, S., Wang, Z. J., and Liao, R. (2016). A cnn regression approach for real-time 2d/3d registration. *IEEE transactions on medical imaging*, 35(5):1352–1363.

[50] Ourselin, S., Roche, A., Subsol, G., Pennec, X., and Ayache, N. (2001). Reconstructing a 3d structure from serial histological sections. *Image and vision computing*, 19(1):25–31.

[51] Papp, E. A., Leergaard, T. B., Csucs, G., and Bjaalie, J. G. (2016). Brain-wide mapping of axonal connections: workflow for automated detection and spatial analysis of labeling in microscopic sections. *Frontiers in neuroinformatics*, 10:11.

[52] Pichat, J., Iglesias, J. E., Yousry, T., Ourselin, S., and Modat, M. (2018). A survey of methods for 3d histology reconstruction. *Medical image analysis*, 46:73–105.

[53] Ren, J., Friedmann, D., Xiong, J., Liu, C. D., DeLoach, K. E., Ran, C., et al. (2018). Anatomical, physiological, and functional heterogeneity of the dorsal raphe serotonin system. *bioRxiv [Preprint]*. Available at: https://doi.org/10.1101/257378.

[54] Rohlfing, T. and Maurer, C. R. (2003). Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*, 7(1):16–25.

[55] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721.

[56] Shamonin, D. P., Bron, E. E., Lelieveldt, B. P., Smits, M., Klein, S., and Staring, M. (2014). Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer's disease. *Frontiers in neuroinformatics*, 7:50.

[57] Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. (2016). A deep metric for multimodal registration. In *International conference on medical image computing and computer-assisted intervention*, pages 10–18. Springer.

[58] Stille, M., Smith, E. J., Crum, W. R., and Modo, M. (2013). 3d reconstruction of 2d fluorescence histology images and registration with in vivo mr images: Application in a rodent stroke model. *Journal of neuroscience methods*, 219(1):27–40.

[59] Tiesinga, P., Bakker, R., Hill, S., and Bjaalie, J. G. (2015). Feeding the human brain model. *Current opinion in neurobiology*, 32:107–114.

[60] Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 13(4):376–380.

[61] Vedaldi, A. and Fulkerson, B. (2010). Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1469–1472. ACM.

[62] Verbeeck, N., Yang, J., De Moor, B., Caprioli, R. M., Waelkens, E., and Van de Plas, R. (2014). Automated anatomical interpretation of ion distributions in tissue: linking imaging mass spectrometry to curated atlases. *Analytical chemistry*, 86(18):8974–8982.

[63] Wachowiak, M. P., Smolíková, R., Zheng, Y., Zurada, J. M., and Elmaghraby, A. S. (2004). An approach to multimodal biomedical image registration utilizing particle swarm optimization. *IEEE Transactions on evolutionary computation*, 8(3):289–301.

[64] Weissbourd, B., Ren, J., DeLoach, K. E., Guenthner, C. J., Miyamichi, K., and Luo, L. (2014). Presynaptic partners of dorsal raphe serotonergic and gabaergic neurons. *Neuron*, 83(3):645–662.

[65] Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., and Shen, D. (2013). Unsupervised deep feature learning for deformable registration of mr brain images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–656. Springer.

[66] Xiong, J., Ren, J., Luo, L., and Horowitz, M. (2018). Mapping mouse brain slice sequence to a reference brain without 3d reconstruction. *bioRxiv [Preprint]*. Available at: https://doi.org/10.1101/357475.

[67] Yang, X., Kwitt, R., and Niethammer, M. (2016). Fast predictive image registration. In *Deep Learning and Data Labeling for Medical Applications*, pages 48–57. Springer.

[68] Yang, Z., Richards, K., Kurniawan, N. D., Petrou, S., and Reutens, D. C. (2012). Mri-guided volume reconstruction of mouse brain from histological sections. *Journal of neuroscience methods*, 211(2):210–217.

[69] Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv [preprint]*. Available at: https://arxiv.org/abs/1511.07122.

[70] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.