MULTIMODAL MRI ANALYSIS
USING DEEP LEARNING METHODS

by
Muhan Shao

A dissertation submitted to The Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
September, 2022

Abstract

Magnetic resonance imaging (MRI) has been widely used in scientific and clinical research. It is a non-invasive medical imaging technique that reveals anatomical structures and provides useful information for investigators to explore aging and pathological processes. Different MR modalities offer different useful properties. Automatic MRI analysis algorithms have been developed to address problems in many applications such as classification, segmentation, and disease diagnosis. Segmentation and labeling algorithms applied to brain MRIs enable evaluations of the volumetric changes of specific structures in neurodegenerative diseases. Reconstruction of fiber orientations using diffusion MRI is beneficial to obtain better understanding of the underlying structures.

In this thesis, we focused on development of deep learning methods for MRI analysis using different image modalities. Specifically, we applied deep learning techniques on different applications, including segmentation of brain structures and reconstruction of tongue muscle fiber orientations. For segmentation of brain structures, we developed an end-to-end deep learning algorithm for ventricle parcellation of brains with ventriculomegaly using T1-w MR images. The deep network provides robust and accurate segmentation results in subjects with high variability in ventricle shapes and sizes. We developed another deep learning method to automatically parcellate the thalamus into a set of thalamic nuclei using T1-w MRI and features from diffusion MRI. The algorithm incorporates a harmonization step to make the network adapt to input images with different contrasts.

We also studied the strains associated with tongue muscles during speech production using multiple MRI modalities. To enable this study, we first developed a deep network to reconstruct crossing tongue muscle fiber orientations using diffusion MRI. The network was specifically designed for the human tongue and accounted for the orthogonality property of the tongue muscles. Next, we proposed a comprehensive pipeline to analyze the strains associated with tongue muscle fiber orientations during speech using diffusion MRI, and tagged and cine MRI. The proposed pipeline provides a solution to analyze the cooperation between muscle groups during speech production.

Thesis Readers

Dr. Jerry L. Prince (Primary Advisor)

Dr. Archana Venkataraman (Second Reader)

Dr. John Goutsias

Acknowledgements

I am very grateful to the exciting journey and memorable mentorship at the Johns Hopkins University and the Image Analysis and Communications Lab (IACL). Working at a big lab, I am deeply indebted to a wonderful group of people, who are not only brilliant scientists in their respective field, but also guide me through this path and provide tremendous help both in my research and real life.

First, I would like to express my deepest gratitude to my advisor, Dr. Jerry L. Prince, for his patient guidance and support in the past six years. It has been a great fortune for me to study under his supervision during my Ph.D. career. His knowledge and passion motivated me to become a better researcher and directed me to the amazing field of medical image study. I would also like to express my appreciation to Dr. Archana Venkataraman and Dr. John Goutsias for being on my thesis committee and providing valuable feedback on my thesis.

I would also like to thank many collaborators across the Johns Hopkins University, the Johns Hopkins Hospital, University of Maryland, Baltimore, and at other institutions. Over the years, my time has been spread across multiple projects, which gives me great opportunities to collaborate with many different brilliant researchers. I would like to thank Dr. Lotta Ellingsen for her patient guidance on my first project after I joined IACL. She also helped me find my bearings as a fresh Ph.D. student. I would like to thank Dr. Ari Blitz and Dr. Dzung Pham for the collaboration in analysis of patient data with hydrocephalus. I would like to thank Dr. Maureen Stone for her support and her

Contents

List of Tables

List of Figures

xvii

Chapter 1

Introduction

1.1 Magnetic Resonance Imaging Modalities and Analysis

Magnetic resonance imaging (MRI) is a non-invasive medical imaging technique that can reveal anatomical structures. It has been widely used to address problems in scientific and clinical research. Numerous automatic MRI processing algorithms have been developed to explore aging and pathological processes. For example, identifying a brain structure from MRI is essential to visualize and track volumetric changes that may occur due to normal aging or neurodegenerative diseases. Reconstruction of fiber tracts is beneficial to obtain a better understanding of the underlying structures. In this dissertation, we are interested in analyzing MRI images of the brain and the tongue. Particularly, we developed algorithms for segmentation of brain structures and reconstruction of tongue muscle fiber orientations and strain analysis. The presented algorithms use images produced by different MRI pulse sequences.

MRI is based on the magnetization properties of atomic nuclei. The contrast of the MRI is mainly determined by the tissue properties and scanning parameters such as

repetition time (TR) and echo time (TE). Different MR modalities can emphasize different tissue types and therefore provide different useful properties. The most commonly used structural MRI sequences include T1-weighted (T1-w) and T2-weighted (T2-w) scans. T1-w MR images are acquired using short TE and TR. T2-w MR images are acquired using longer TE and TR. The structural MRI sequences can be acquired at high resolution and provide good contrast in the human brain. The diffusion weighted imaging (DWI) generate signal contrast based on the differences in Brownian motion of water molecules in the biological tissue. Therefore, DWI can quantify the anisotropy of diffusion and analyze tissues with highly organized fiber structures. Some measurements can be derived from DWI to characterize the diffusion property of the underlying tissue, providing additional contrast that is not visible in standard structure MRI. It is common in clinical and research study to acquire multiple modalities of MR images to reveal different information about the region of interest. Figure 1-1 shows examples of different MRI modalities used in this dissertation to study the human brain. Besides revealing anatomical structures, MRI can also be used to capture tissue motion. The internal motion of the tissue can be measured by tagged and cine MR sequences. Theses scans are acquired during motion over time. Motion information can be reconstructed from the deformed magnetic tag patterns. Details of the tagged and cine MR imaging can be seen in Section 5.1.2. Figure 1-2 shows example of different MRI modalities used in this dissertation to study the human tongue.

Many widely-used MRI analysis software packages [1, 2] rely on only one MRI sequence, typically T1-w MRI, for brain structure segmentation. However, it is difficult to differentiate the boundaries of some structures of interest, such as the thalamic nuclei, using only T1-w MRI. Other available MRI modalities can provide properties to improve the segmentation results of specific structures. In the research on tongue muscle cooperation pattern, many existing studies [3, 4] only utilize tagged MRI to analyze tongue motion. There is very little work on the joint use of multimodal MRIs to

2

Axial	Coronal	Sagittal

T1-w MPRAGE

DWI, b = 0

DWI, b = 1000
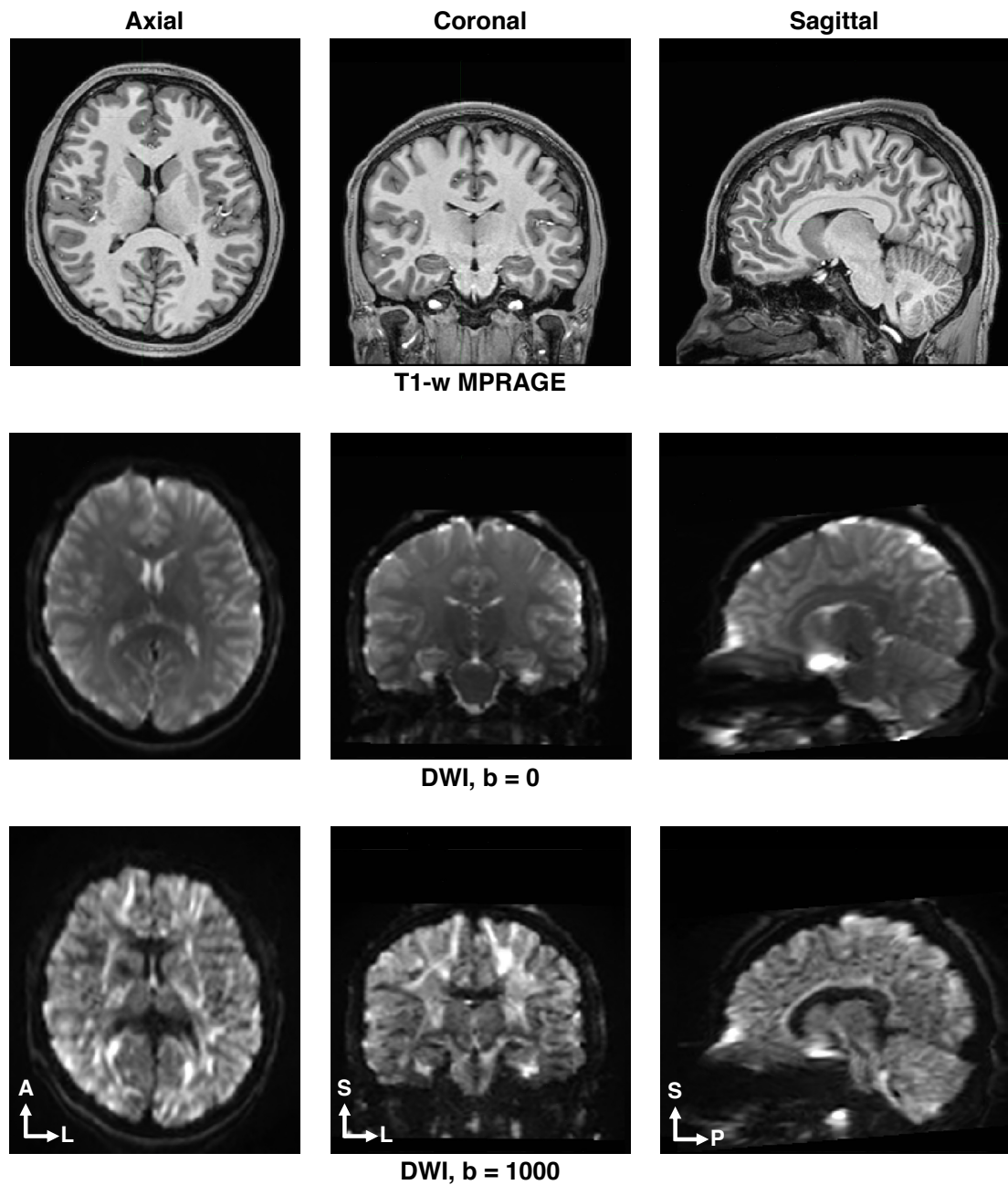
Figure 1-1. Examples of different MRI modalities of the same subject at the brain. A = anterior; L = left; S = superior; P = posterior.

Axial	Coronal	Sagittal

T2-w

Cine

DWI, b = 0

DWI, b = 500
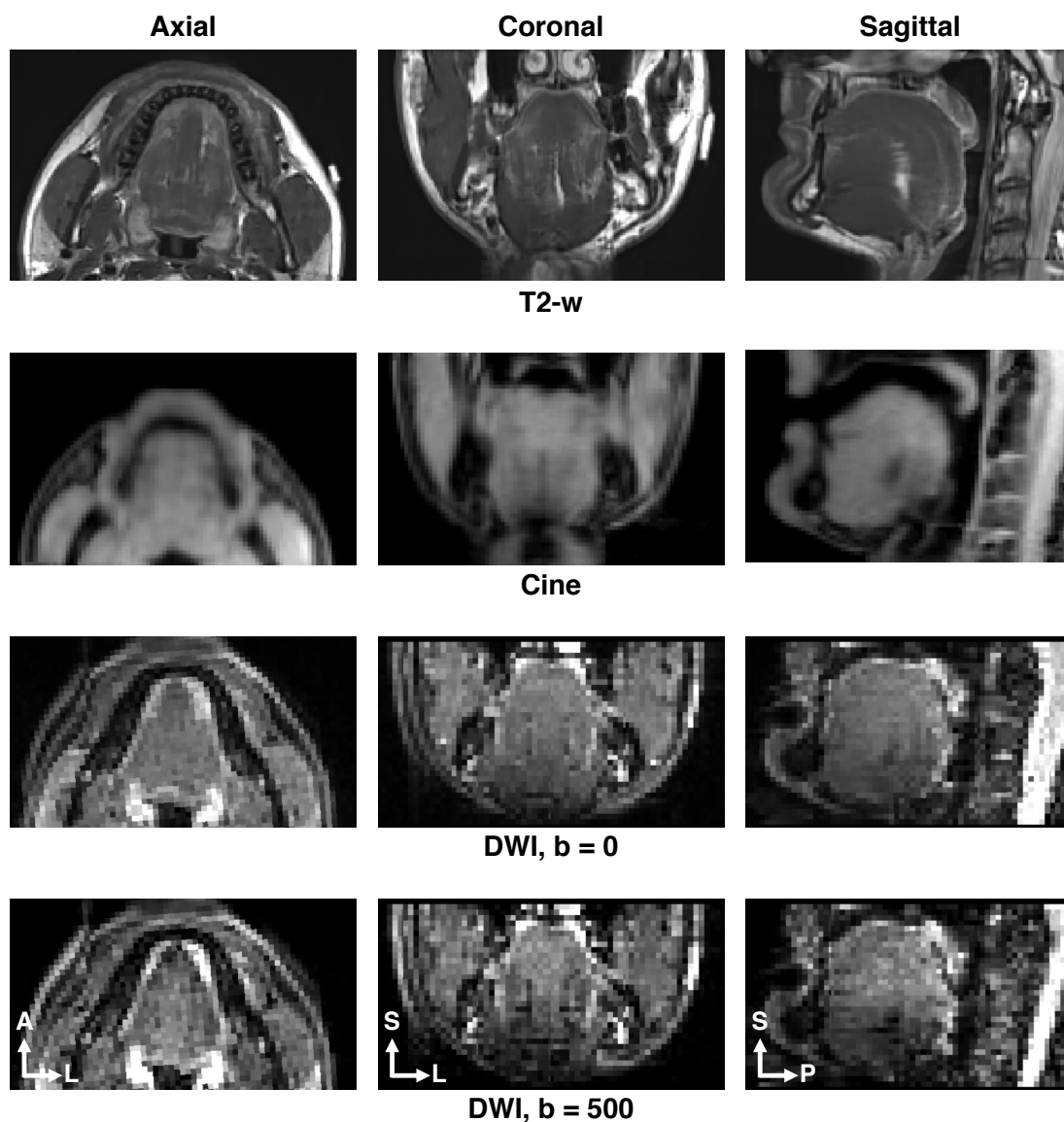
Figure 1-2. Examples of different MRI modalities from the same tongue data. A = anterior; L = left; S = superior; P = posterior.

acquire accurate estimation of muscle fibers and tongue motion. In this dissertation, we present specialized algorithms to analyze MR images using multiple pulse sequences.

1.2 Deep Learning in MRI Analysis

Deep learning is an advanced machine learning technique that has been widely applied in medical image analysis in recent years. It is proving to achieve enhanced performance over conventional methods in various medical applications [5, 6]. In this dissertation, we focus on using deep learning techniques to develop MR image analysis algorithms. In this section, we provide a brief introduction to the deep learning techniques used in this dissertation.

Deep learning techniques use deep artificial neural networks to extract informative features from the input data. A typical deep network is composed of many layers that transform the input (e.g., the MR images) to output (e.g., disease classification or structure segmentation) by learning hierarchical feature representations. Each layer is a container that usually receives an input feature map, transforms it with a linear or non-linear operation, and then passes it as output to the next layer. Most commonly used linear operations include fully connected layers and convolutional layers. Non-linear operations, also called activation layers, include the sigmoid function and the rectified linear activation function [7]. Other commonly used layers include downsampling (pooling) and upsampling layers, dropout layers [8], and feature normalization layers [9, 10]. Non-linear operations increase the complexity of the relationships between the input and output of the network. A sufficient number of linear and non-linear layers enable a network to approximate any underlying function in mapping the inputs to the outputs [11], which is parameterized by the model weights (trainable parameters).

Like conventional machine learning techniques, deep learning requires data to learn the model parameters, which can be formed as minimizing a predefined loss function. This loss function is a measurement that evaluates the error between the model prediction and the truth. For example, mean squared error and cross-entropy error are commonly used in deep network training. Typically, there is no analytic solution

for the model parameters to minimize the loss function. Therefore, stochastic gradient descent (SGD) is often used to optimize the loss function. The gradient of the model parameters are calculated via the back-propagation algorithm [12, 13]. The model parameters are updated iteratively until convergence. During each update iteration, instead of computing the derivative over the whole training set, SGD randomly samples a small subset of training data, called a mini-batch, to compute the derivative, which greatly reduces the computational complexity. An adaptive version of SGD, called Adam [14], introduced adaptive learning rates for different parameters. It shows fast convergence, and the hyper-parameters are easy to interpret and require little tuning. Adam has been widely used in modern deep learning techniques.

Depending on the availability of ground truth "label" data during training, deep learning can be supervised or unsupervised. In typical supervised deep learning, the training process have labeled data to help make predictions. A segmentation network with the labeled mask and an image synthesis network with the target image are two examples of supervised deep learning. Unsupervised learning does not have labeled data and must discover the inherent patterns within the training data. Generative models such as a variational autoencoder [15] and a generative adversarial network [16] are characterized as unsupervised deep learning. In this dissertation, we focus on supervised deep learning for MRI analysis.

A big challenge in training deep network for medical image analysis is overfitting, which mostly arises from limited training data compared to the model complexity. Several strategies have been proposed to improve the generalizability of deep networks. Data augmentation [17] introduces small variations to the original data to increase the amount of training data. Dropout [8] is a regularization method that helps prevent the neural network from overfitting. It randomly "ignores" some nodes during training, therefore encourage the nodes within a layer to learn more representative information. Another type of methods to reduce generalization error is to standardize the inputs to a layer

using methods such as batch normalization [9] and instance normalization [10]. These techniques have the effect to stabilize the network training process and greatly accelerate its convergence.

The convolutional neural network (CNN) is one of the most popular deep learning architectures in medical image analysis. In CNNs, the model weights are shared by performing a multiplication between the input array and a small kernel, which makes it shift-invariant and also drastically reduces the amount of the trainable parameters. A convolutional layer is typically followed by a normalization layer and then an activation layer. The CNNs often incorporate pooling layers to downsample the feature map, which increase the receptive field of the subsequent convolutional layers and enable the network to extract features from the input image hierarchically. In this dissertation, we explore the CNN technique on segmentation of brain structures and reconstruction of tongue muscle fiber orientation from multimodal MR images.

U-Net [18, 19] is one of the most popular deep network architectures in medical image segmentation. It is evolved from the conventional convolutional neural network. Figure 1-3 shows a general architecture of U-Net. The name of U-Net comes from its "U" shape. This architecture consists of two major parts: the contracting path on the left and the expanding path on the right. The contracting path consists of encoder blocks that downsample the input feature image to capture high-level information. Then the decoder blocks in the expanding path restore the image by upsampling. The long skip connections (see green arrows in Figure 1-3) combine the high resolution features from the contracting path with the upsampled features from the expanding path to help recover the full spatial resolution at the output. U-Net shows superior performance in applications with small amount of training data, which is typically the case in medical image segmentation.
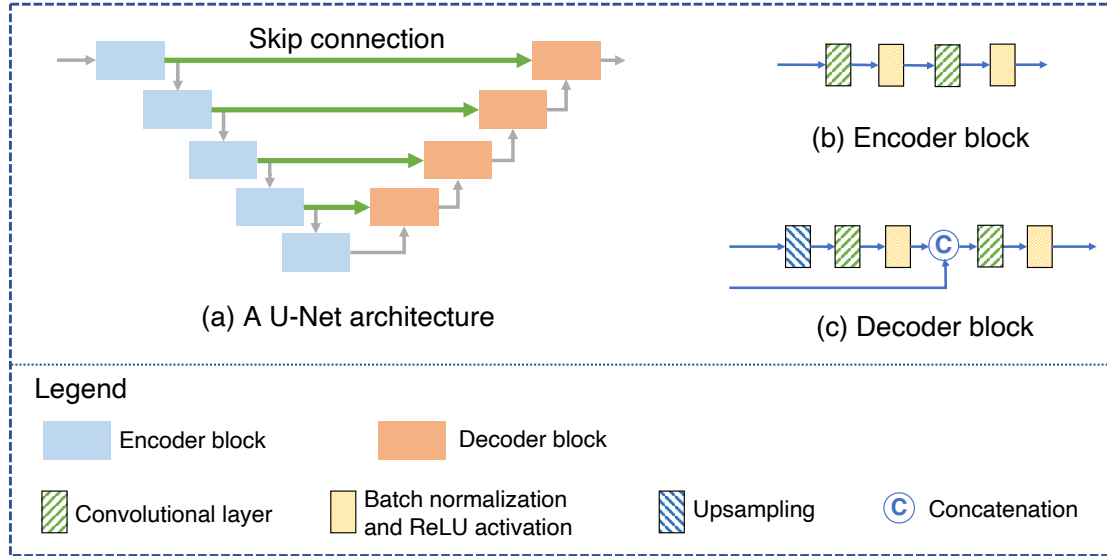
Figure 1-3. (a) An example of U-Net architecture. The green arrows represent the skip connections. (b) Encoder block. (c) Decoder block.

1.3 Dissertation Overview

1.3.1 Contributions

In this thesis, we developed a set of deep learning methods for MRI analysis using different image modalities. The four main contributions of this research are summarized as follows.

Parcellation of the Brain Ventricles

The first contribution is the development of a CNN-based method [20, 21] to parcellate the brain ventricular system into its sub-compartments using T1-w MRI. The ventricular system of the human brain is a communicating network consisting of cavities filled with cerebrospinal fluid (CSF). Disruption of the CSF flow could lead to enlarged ventricles and cause brain damage. Studies have shown that the identification of disproportionate expansion of the ventricular system could help better understand some neurodegenera-

tive diseases [22, 23]. This leads to the motivation behind this work since there is little work on automatic segmentation tool for parcellation of the ventricular system in patients with severely enlarged ventricles. The proposed method, called ventricle parcellation network (VParNet), is modified from a 3D U-Net and incorporates residual units in the network. It has been demonstrated to achieve better accuracy in shorter computational time compared to widely used brain segmentation methods in data sets with highly variable ventricle sizes and shapes. VParNet provides a tool for automated parcellation of the ventricular system in healthy subjects and patients with ventriculomegaly (enlarged ventricles), enabling more sophisticated analyses of the ventricular structures in both health and disease. It is publicly available as a Singularity container at http://iacl. ece.jhu.edu/index.php?title=Brain_ventricle_parcellation_instructions.

Parcellation of the Brain Thalamus

The second contribution is the development of a multi-channel 3D U-Net method to segment the brain thalamus and parcellate each hemisphere into thalamic nuclei using T1-w and diffusion MRIs [24]. The thalamus is a gray matter subcortical structure that acts as a relay station for the human brain. It can be subdivided into thalamic nuclei based on histological and functional criteria. Volume changes in thalamic nuclei are of great interest in many neurodegenerative studies [25–27]. It is essential to have an accurate and consistent segmentation of the thalamus to better manage the progression of the diseases. Our proposed network takes T1-w intensity image and two feature images derived from diffusion MRIs as input. We incorporated an MR image harmonization algorithm [28] into the data pre-processing. The proposed method achieved better accuracy on the thalamus segmentation compared to widely used segmentation algorithms. The image harmonization step adjusts the image contrast while maintaining the underlying anatomy. With this technique, the network provides the potential to process large amounts of data from different data sets, even in the absence

of site-specific training data.

Reconstruction of Tongue Muscle Fiber Orientations

In the third contribution, we developed a deep learning method to reconstruct the crossing muscle fiber orientations in the human tongue using diffusion MRI [29]. The tongue plays an important role in multiple vital human functions including breathing, swallowing, and speaking. These functions are possible because of the complex muscle architecture of the tongue. There are compelling scientific reasons to obtain a better understanding of the relationship between individual tongue muscles and the complex functions [30]. Therefore, it is important to obtain an accurate reconstruction of the tongue muscle fiber orientations. Most existing methods were designed for brain tissue and do not account for the orthogonal nature of the crossing muscle fibers in the human tongue. We proposed a 3D patch-based CNN algorithm to reconstruct the tongue muscle fiber orientations from diffusion MRI. We apply a spherical harmonic transformation to the input image to make the network adapt to different imaging protocols. We proposed a separation loss in the training step to account for the orthogonality of the crossing muscle fibers in the tongue, which is different from white matter tracts in the brain. This method outperforms conventional fiber orientation estimation algorithms on synthetic tongue data.

Another algorithm was developed to further refine the reconstructed fiber orientations. The crossing fiber orientations produced by the reconstruction algorithms sometimes show inconsistent assignment of the first and second fiber orientations. Therefore, we proposed a fiber matching algorithm [31] to refine the assignment of the estimated fiber orientations. A quadratic unconstrained binary optimization framework was formed to find an optimal assignment for each voxel in the tongue region. It is demonstrated to improve the visualization of the fiber orientations in both synthetic and real tongue data. The proposed fiber matching algorithm can also be used as a post-processing step for

other fiber orientation reconstruction algorithms.

Analysis of Strain in the Line of Action

In the fourth contribution, we developed a pipeline to analyze the strain in the line of action along tongue muscle fiber orientations during speech production. In speech production, the tongue helps to form sounds via highly complex deformation patterns. Analyzing the interaction and cooperation between tongue muscles is of great scientific and clinical interest in the study of speech production [32, 33]. It is also important to understand the adaptive behaviors of the tongue in patients with glossectomies, where the tongue has been partly surgical removed. There is very little work on the joint analysis of the muscle fibers and the tongue motion, which are both required to analyze the muscle cooperation pattern in speech production. Our proposed pipeline jointly analyze diffusion MRI and tagged and cine MRI to reveal strain components in the tongue muscle fiber orientations during speech. Diffusion MRI was used to obtain the muscle fiber orientations, as described in the third contribution. Tagged and cine MRIs were used to obtain the tongue motion during speech [34]. Strains in the line of action are calculated by projecting the computed strain tensors onto local muscle fiber orientations. We also adopted a manually labeled tongue muscle mask to analyze the strain pattern within individual muscle groups. We performed this analysis on a cohort of 8 healthy controls and 2 post-partial-glossectomy patients and achieved both qualitative and quantitative results that showed some common patterns in the controls and differences in the patients. This new pipeline has the potential to aid researchers in better understanding the cooperation between muscle groups in speech production and in learning about the adaptive behaviors of patients with glossectomies.

1.3.2 Organization

The remainder of the dissertation is organized as follows. In Chapter 2, we describe the deep learning algorithm for ventricle parcellation using a T1-w MR image. In Chapter 3, we present the deep learning algorithm for thalamus parcellation using a T1-w MR image and DTI features. In Chapter 4, we present the tongue muscle fiber orientation reconstruction network and the fiber orientation matching algorithm. In Chapter 5, we describe the pipeline for strain analysis during speech production. In each chapter, we provide background on the project, describe the image processing algorithm, and evaluate the algorithm using a set of experiments. Finally, Chapter 6 includes a summary of the contributions and presents future research directions.

Chapter 2

Brain Ventricle Parcellation using T1-w MRI

2.1 Introduction

2.1.1 Background

The ventricular system of the human brain consists of four cavities: two large lateral ventricles and the third and fourth ventricles (see Figure 2-1(a)). All four ventricles contain strands of a highly convoluted and vascular membranous material called choroid plexus, which is a network of ependymal cells involved in the production of cerebrospinal fluid (CSF). The CSF fills the ventricular space within the brain and bathes the entire central nervous system. The entire volume of CSF is renewed two to three times per day [35, 36] and disruption of the flow can cause excess CSF to build up leading to the clinical condition called hydrocephalus. In hydrocephalus, the ventricles expand and press against nearby brain tissue causing the brain shape to become distorted (see Figure 2-1(b)), eventually leading to brain damage. If this happens over a long period of time, the clinical syndrome of normal pressure hydrocephalus (NPH) may

result. In conjunction with ventriculomegaly, typical symptoms of NPH include cognitive impairment, gait dysfunction, and urinary incontinence [37].
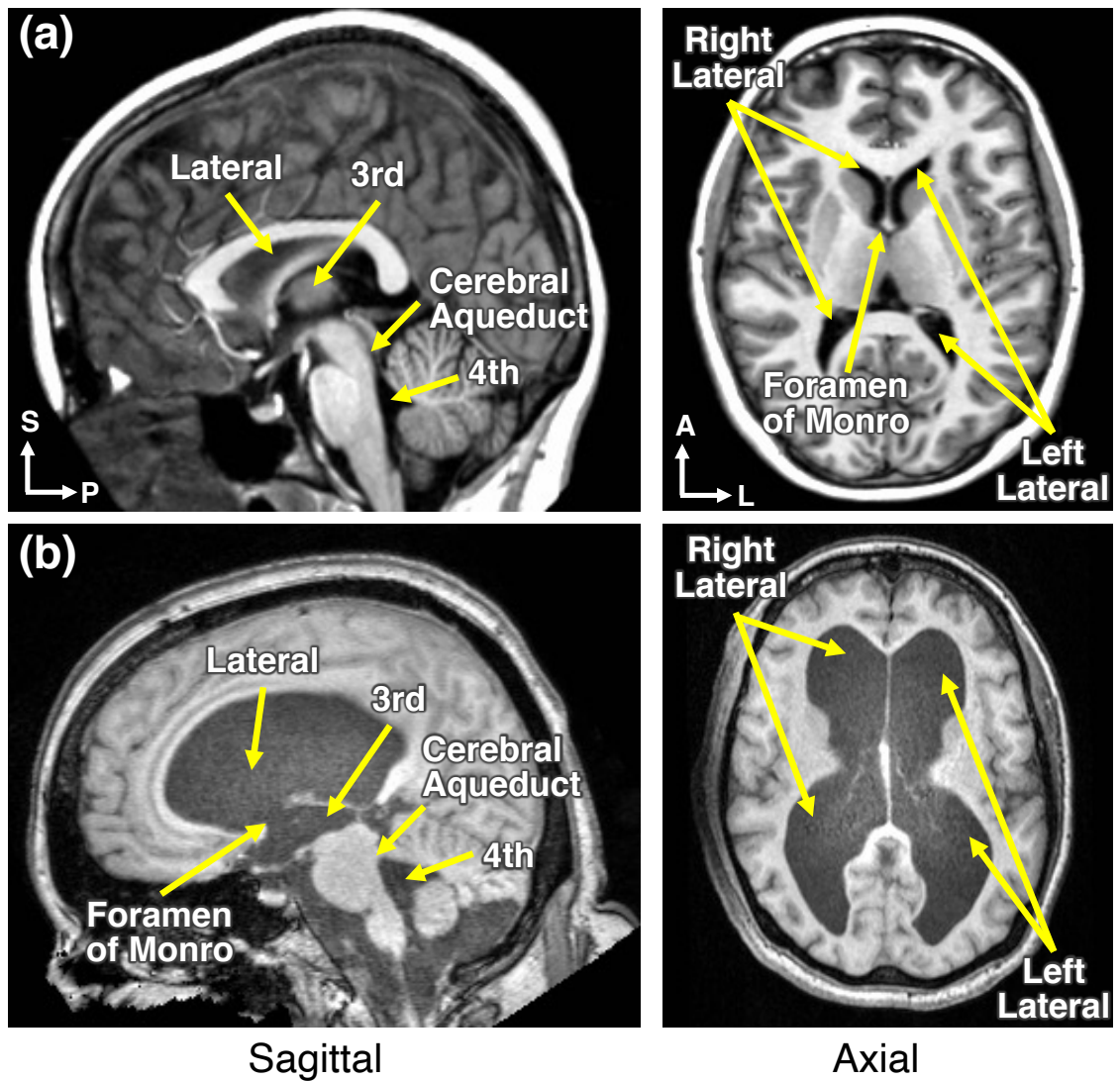
Figure 2-1. The ventricular system on an MPRAGE T1-w MRI of (a): a healthy subject and (b) an NPH subject. For each subject, the left image shows a sagittal slice and the right image shows an axial slice. S: superior; P: posterior; A: anterior; L: left.

NPH most often affects older adults and may account for more than five percent of all dementia diagnoses [38]. However, unlike other well-known causes of dementia, such as Parkinson's and Alzheimer's disease, once diagnosed, NPH patients have the option of shunt surgery or endoscopic third ventriculostomy (ETV), making the dementia-like

symptoms, as well as other symptoms of the disease, potentially reversible. Recent studies suggest that the prevalence of NPH is much higher than the number of persons actually treated [39] and given this existing therapy, it is important to diagnose those patients who will respond well to ETV or shunt surgery [40]. However, the diagnosis remains challenging since there are no distinctive pathognomonic features for idiopathic NPH and symptoms overlap with other dementias [40–42]. It has even been suggested that the definitive diagnosis of this condition should be based on the patient's response to the surgery itself [43].

At present, NPH is diagnosed by physical examination (gait evaluation) and computed tomography (CT) or magnetic resonance imaging (MRI) of the brain, and the likelihood of a positive response to therapy is determined using a lumbar puncture [43]. Patients who show clinical improvement (as measured by improvement in gait) following removal of some CSF by lumbar puncture become candidates for shunt placement [44] or ETV. However, lumbar puncture has shown to have very low negative predictive value (< 20%), as a number of patients improve after a shunt-placement despite a negative lumbar puncture test [45]. Hence, a more sensitive diagnostic method is needed to identify more treatment-responsive NPH patients.

Given the constant production of CSF, it is critical that its circulation remains uninterrupted. The circulation of CSF can be obstructed at any point in the flow pathway, although the foramina of Monro (which connect each of the lateral ventricles to the third ventricle) and the cerebral aqueduct (connecting the third and fourth ventricles) are well-known anatomic bottlenecks because they are intrinsically predisposed to blockage [35]. An obstruction in either or both of the foramina of Monro causes the corresponding lateral ventricle to expand and become hydrocephalic, without distorting the remainder of the ventricular system. On the other hand, an obstruction in the cerebral aqueduct causes expansion of the third ventricle and both lateral ventricles [35]. Disproportionate expansion of components of the ventricular system could therefore give valuable clues

about the specific point of CSF obstruction, which could in turn have an impact on whether ETV or shunt surgery will be effective [22, 23]. This line of reasoning leads to the motivation behind our work since accurate automated segmentation and labeling tools for the multiple different compartments of the ventricular system of patients with severely enlarged ventricles was not present prior to our work.

2.1.2 Related Work

Labeling the ventricular system as one component in healthy subjects using MRI is carried out routinely by several algorithms and their associated software packages. Although labeling the ventricular system in NPH patients seems like a simple task, many of these methods fail when applied to NPH brains [46, 47]. The major challenge when parcellating and labeling brain with severe ventriculomegaly is that the ventricles in these patients are both greatly enlarged and distorted from their normal anatomical shape, which makes accurate labeling of the major compartments challenging (see for example the lateral and third ventricles in Figure 2-1(b), left image). Furthermore, a key limitation of these methods is their inability to parcellate the ventricular system into multiple compartments in order to quantify disproportionate dilation of the different ventricular cavities.

Parcellation and labeling of the ventricular system into its four main compartments (i.e., left and right lateral and third and fourth ventricles) has also been carried out by several approaches. Table 2-I presents an overview of these segmentation algorithms. Multi-atlas label fusion methods have been shown to outperform other methods in the task of parcellating multiple brain structures, both on healthy and several (non-NPH) diseased populations [64]. However, these methods have a long runtime (several hours) and tend to fail in cases with enlarged ventricles, as demonstrated in Figure 2-2. One reason for some of these failures is that most methods rely on registration between the

Table 2-I. Overview of brain segmentation methods

Method	Whole ventricle label	Ventricle parcellation	Remarks
volBrain [53]	✓		Non-local patch-based label fusion method. Provides online MRI brain volumetry system.
ALVIN [54]	✓		Applies a binary mask to CSF segmented images using "unified segmentation" in SPM8 to segment the lateral ventricles.
TOADS [55]	✓		Segmentation framework based on both topological and statistical atlases of brain anatomy.
LoAD [56]	✓		Model-based segmentation method providing post refinements to a probabilistic segmentation model with anatomical tissue priors.
Adaptive Atlases [46]	✓		Generates a subject specific atlas to segment brains with ventriculomegaly.
S3DL [47]	✓		Patch-based segmentation method using sparse dictionary learning.
FreeSurfer [2, 48, 49]	✓	✓	Atlas-based approach for whole brain segmentation.
MUSE [57]	✓	✓	Multi-atlas label fusion method integrating optimal atlas selection strategy and a boundary modulation term to refine the segmentation.
BrainSuite [58]	✓	✓	Atlas-based method for brain surface and volume labeling.
MALPEM [50]	✓	✓	Multi-atlas label fusion method using a relaxation scheme to correct registration error.
NLSS [59]	✓	✓	Multi-atlas segmentation method with statistical fusion to incorporate intensity into the estimation process.
Joint label fusion [52, 60]	✓	✓	Multi-atlas label fusion method formulating a weighted voting scheme to minimize the total expectation of the labeling error.
RUDOLPH [61–63]	✓	✓	Combines tissue segmentation and multi-atlas segmentation to correct registration priors. Designed for subjects with ventriculomegaly.

| NPH subject | FreeSurfer | MALPEM | Joint Label Fusion |
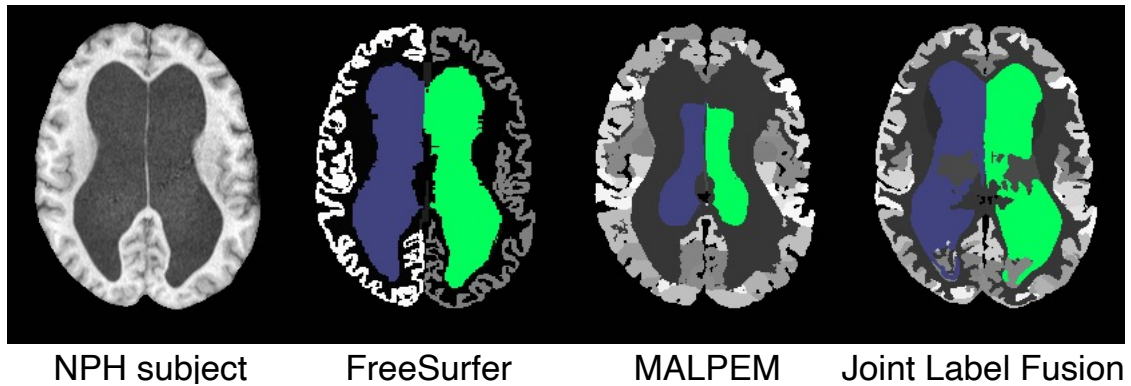
Figure 2-2. An example of a failed segmentation on a subject with severe ventriculomegaly, due to NPH. The segmentation results were produced by three algorithms: FreeSurfer [2, 48, 49], MALPEM [50], and Joint Label Fusion from the ANTs software package [51, 52].

atlas and subject images, which in pathological cases is rarely optimal. One algorithm, called RUDOLPH [61–63], was specifically designed to segment subjects with enlarged ventricles. Although robust in ventricle parcellation, this method takes hours to process a single image.

2.1.3 Contribution

In this chapter, we present a CNN-based method to automatically parcellate the brain ventricles into its sub-compartments from T1-w MRI [20, 21]. The method was modified from 3D U-Net [19] and can be applied to both healthy subjects and patients with ventriculomegaly. The method was evaluated by comparing with other state-of-the-art algorithms against manually delineated ground truth. Finally, a pilot study was carried out, examining the performance of the proposed method on a patient data set with implanted valves in the brain. The proposed network is publicly available as Singularity container at http://iacl.ece.jhu.edu/index.php?title=Brain_ventricle_parcellation_instructions.

2.2 Methods

2.2.1 Data Acquisition and Pre-processing

Images from two separate data sets of T1-weighted (T1-w) magnetically prepared rapidly acquired gradient echo (MPRAGE) [65] MRI scans were used to train and evaluate our method. The first data set comprises 50 MRIs of healthy controls (age range: 18–96 years with mean age of 44.54 years and standard deviation of 25.16 years) from the Open Access Series on Imaging Studies (OASIS) data set [66]. They were acquired on a 1.5T Siemens scanner with scanner parameters of: TR = 9.7 ms, TE = 4.0 ms, FA = 10°, TI = 20 ms, and $1 \times 1 \times 1.25$ mm^3 voxel size. Written informed consent was obtained from all participants. The images were manually labeled by experts from Neuromorphometrics Inc (NMM)[1]. The labeled images comprise 134 cortical and subcortical structures, including the left and right lateral ventricles, and the third and fourth ventricles. Since our algorithm focuses on ventricle parcellation, we converted the manual labels into five: the four ventricle labels and a background label representing the remaining parts of the image.

The second data set contains 95 T1-w MPRAGE MRIs from our own NPH database (age range: 26–90 years with mean age of 66.83 years and standard deviation of 15.08 years). They were acquired on a 3T Siemens scanner with typical parameters (with small variations): TR = 2110 ms, TE = 3.24 ms, FA = 8°, TI = 1100 ms, and $0.859 \times 0.859 \times 0.900$ mm^3 voxel size. The acquired data was retrospectively included in our study with local institutional review board approval. The four ventricles of the 95 NPH images were manually delineated. This was done by first identifying the anatomical structure of the ventricles, which required 3–4 hours per patient. These were reviewed by separate experts in neuroanatomy, with either minor editing or returned to the delineator for correction. Once a ventricular system mask was agreed, the components of right

---

[1]http://www.neuromorphometrics.com/

and left lateral ventricles, both foramina of Monro, third ventricle, cerebral aqueduct, and fourth ventricle were identified. This manual parcellation of the ventricular system took another hour per patient to complete. The cerebral aqueduct and the foramina of Monro are not included within our testing, as there do not currently exist such detailed anatomical atlases of the ventricular system in use elsewhere. Thus for evaluation purposes, the foramina of Monro is included with the corresponding lateral ventricle, and the cerebral aqueduct with the fourth ventricle, making the labeling comparable with NMM.

The first step of the pre-processing is intensity inhomogeneity correction using N4 [67]. In the second step of pre-processing, the inhomogeneity-corrected image is rigidly registratered to MNI 152 atlas space [68]. The third step is skull-stripping [69]. The original image size after all the pre-processing steps is $241 \times 286 \times 241$ voxels; however, each image is symmetrically cropped around the brain mask to $192 \times 256 \times 192$ voxels to reduce the number of background voxels before being input into our deep network.

2.2.2   Ventricle Parcellation Network

Given an MRI of a human brain, our goal is to assign a particular label to each voxel. In this current work, we use a total of five labels to parcellate the T1-w image, and these labels consist of four ventricle labels and a background label. Our network, referred to as the Ventricle Parcellation Network, or VParNet, was modified from [70] and designed to label each voxel with one of the five labels. It consists of a contracting path with a series of encoder blocks and an expanding path with a series of decoder blocks, as illustrated in Figure 2-3. The long skip connections from the encoder blocks provide the high-resolution features to their matching decoder blocks through concatenation [19]. The $1 \times 1 \times 1$ projection convolutions connected to each decoder block reduce the

number of output channels to the number of labels, which is five in our case. The details of the different blocks are described below.

Encoder block

The encoder block in the contracting path, as shown in Figure 2-3(b), is similar to the residual unit in [71]. We refer to the instance normalization layer [10] and leaky rectified linear unit (ReLUx) layer [72] as activation layers. The input feature map of each encoder block goes through a $3 \times 3 \times 3$ convolution layer with stride setting to two, activation layers, a $3 \times 3 \times 3$ convolution layer, and activation layers. The skip connections within the blocks act as identity maps. Their outputs are combined with the outputs of the second convolution layer using element-wise summation. The short skip connections make error reduction faster and increase accuracy [73] during training. Compared to batch normalization [9], instance normalization can make the feature maps invariant to linear transformation of the input image intensities [10].

Decoder block

The decoder block in the expanding path, as shown in Figure 2-3(c), consists of a $1 \times 1 \times 1$ convolution layer, activation layers, an upsampling layer, a $3 \times 3 \times 3$ convolution layer, activation layers, a concatenation layer, a dropout layer, a $3 \times 3 \times 3$ convolution layer, and activation layers. In the upsampling layer, the coarse feature map is upsampled to a finer resolution by repeating the data in three dimensions. The upsampling layer is followed by a convolution and activation layers, whose output is concatenated with the output feature map from the matching encoder block within the concatenation layer. The dropout layer is used to regularize the network [8].
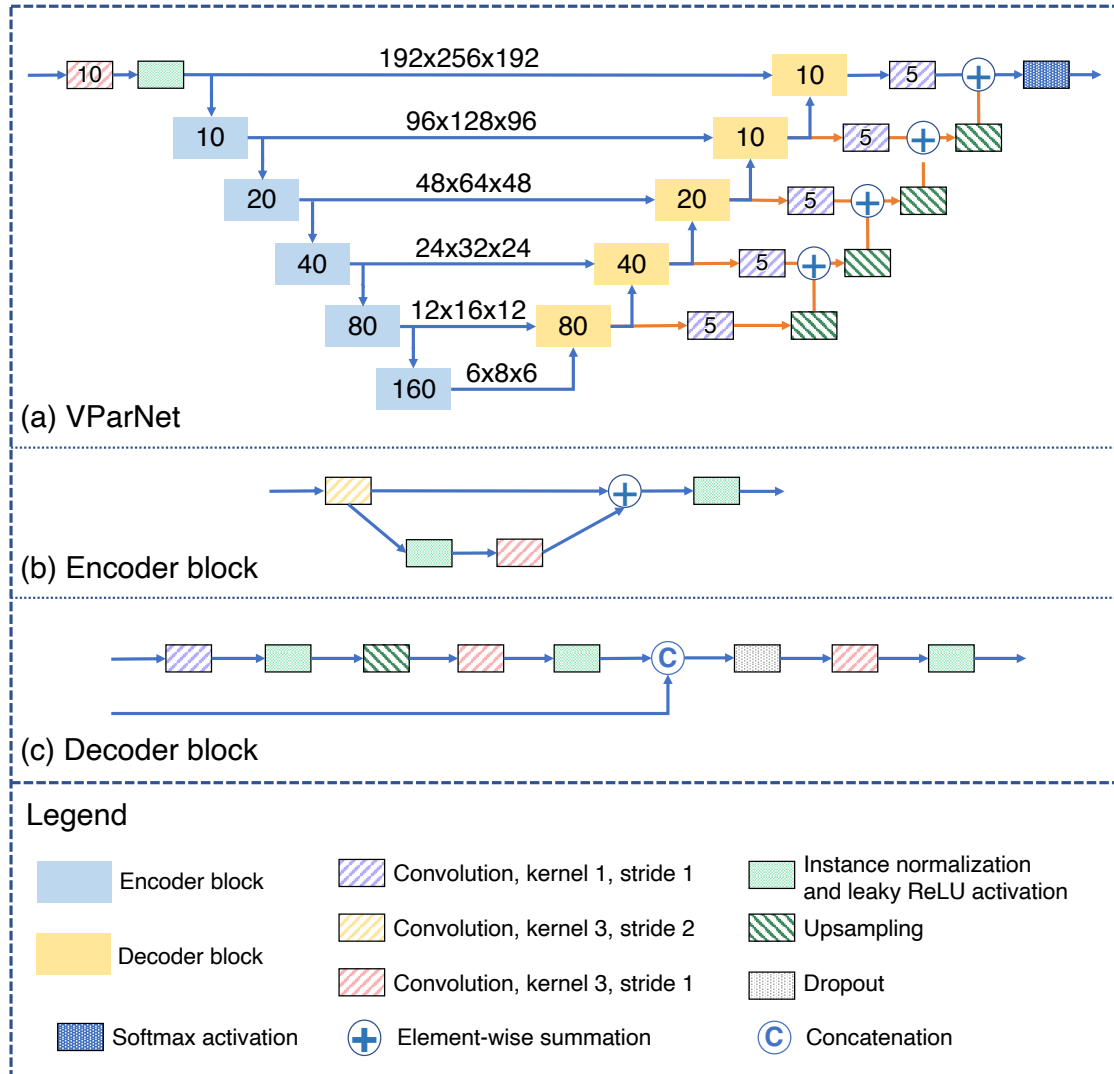
Figure 2-3. Architectures of (a) the ventricle parcellation network (VParNet), (b) the encoder block, and (c) the decoder block. The numbers in the encoder and decoder blocks indicate the number of output channels. The shape of the feature tensor is denoted at each resolution level.

Classification block

The final decoder block is followed by the activation layers and a $1 \times 1 \times 1$ convolution layer to reduce the number of channels to 5, i.e., the number of output labels. The classification result is refined by combining feature maps created from multiple resolution levels [70, 74, 75]. This step encourages the features extracted at different layers to be consistent. The multi-level feature maps are combined in the following way: the feature maps from different levels go through a $1 \times 1 \times 1$ convolution layer to become 5-channel feature maps. Then the 5-channel feature map from the lowest resolution level is upsampled to have the same resolution as the second-lowest feature map. These two are combined via element-wise summation. The combined feature map is upsampled to a higher resolution level and added to the feature map at this resolution level. The upsampling and summation are done sequentially until we get a feature map with the highest resolution, as shown by the orange arrows in Figure 2-3(a). The last summation operation is followed by a softmax activation layer, converting the feature map to a probability map for the final classification.

2.2.3 Data Augmentation

Due to the limited amount of available training data, we augmented the training data to improve the generalization of the network inference (see Figure 2-4 for examples). Specifically, we applied left-right flipping, random rotation, and elastic deformation to the training data. Detailed implementation of each data augmentation method is described below.

- **Left-right flipping** (Figure 2-4(b)). Since human brains are approximately left-right symmetric, we introduced left-right flipping to create more training data. An T1-w MR image and its corresponding manual delineations were flipped left-to-right. The left and right lateral ventricle regions in the flipped label image were relabeled

23
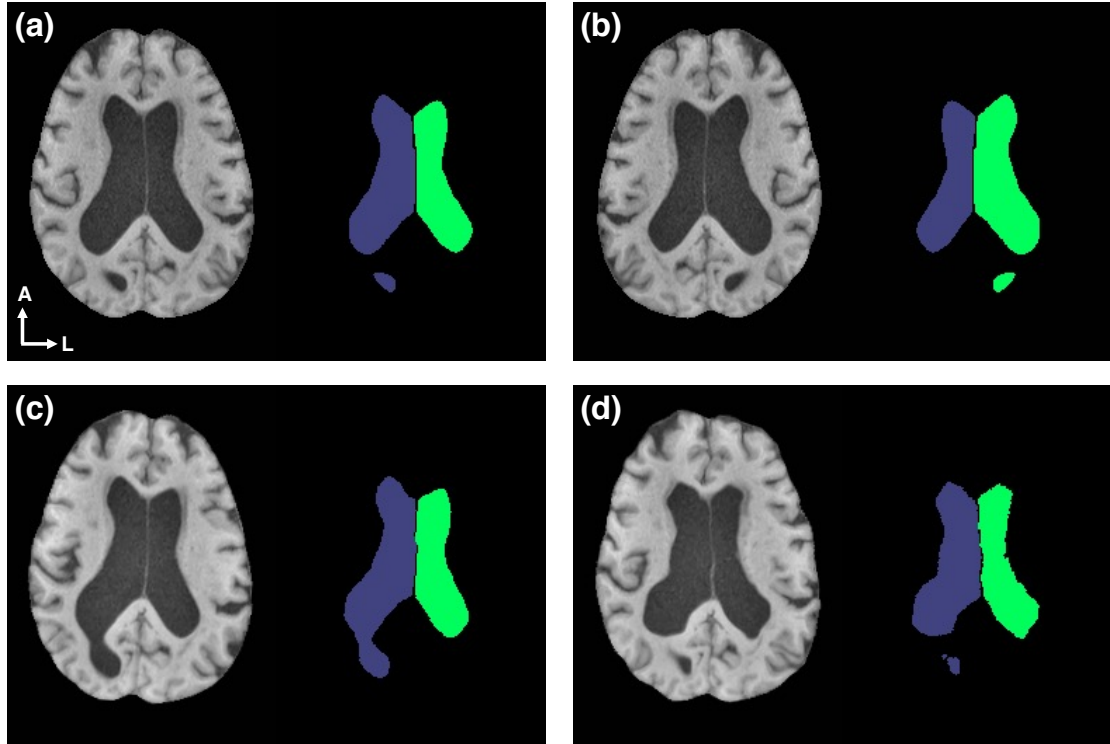
Figure 2-4. Data augmentation examples (MRI and the corresponding label image). (a) The original image; (b) Left-right flipping; (c) Random rotation; (d) Elastic deformation. A = anterior; L = left.

to match their sides.

- **Random rotation** (Figure 2-4(c)). An T1-w image was rotated along the x, y, and z axes with random angles. The random angle along each axis was uniformly sampled from -10 to 10 degrees. The corresponding manual label image was rotated the same way.

- **Elastic deformation** (Figure 2-4(d)). At each voxel in the deformed T1-w image, three random translation values were uniformly sampled from -8 to 8, representing the translation along x, y, and z axes. The new intensity was calculated by interpolation at the translated coordinates in the original image. A Gaussian smoothing with standard deviation equal to 5 voxels was applied separately to the x, y, and z components after the interpolation. The corresponding manual label

image was deformed the same way.

Random rotation and elastic deformation were applied to the left-right flipped images as well as the original images, but these two transformations were not composed.

2.2.4 Network Implementation

The VParNet was trained on images from both the NMM and NPH data sets. Of the 50 MRIs from the NMM data set, 15 were used for training, 5 were used for validation, and the remaining 30 were used for testing. Of the 95 NPH images, 25 were chosen for training, 5 for validation, and the remaining 65 were used for testing. The 95 NPH images were sorted by the volume of the ventricular system and the 25 training MRIs were chosen such that they covered the entire spectrum of ventricle sizes in the NPH cohort.

During network training, the leaky ReLU negative slope was set to 0.1 and the dropout rate in the decoder blocks was set to 0.2. The loss function used to train the network was one minus the mean Dice similarity coefficient (DSC) [76] of all the labels, which is defined as

$$\text{Loss} = 1 - \frac{1}{L} \sum_{l=1}^{L} \frac{\epsilon + 2\sum_i P_{il} T_{il}}{\epsilon + \sum_i P_{il} + \sum_i T_{il}},$$

where $l \in \{1, ..., L\}$ is the label index, L is the total number of labels, P_{il} is the probability that voxel i belongs to the label l generated by the network, and T_{il} is the binary value indicating if voxel i belongs to the label l based on the manual delineations. The ϵ variable is used to avoid a zero denominator, and it was set to 10^{-3} during training.

VParNet was initially trained for 200 epochs using the Adam optimizer [14] with step size $\alpha = 0.001$, exponential decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$. During each epoch, the original 40 images (15 from NMM and 25 from NPH) were first left-right flipped to create another set of 40 images. Then the 80 images were randomly

rotated along each axis and deformed to create an additional 160 images. In each epoch, these 240 (80+160) images were used to optimize the network parameters, with a batch size of one. We observed that the performance of the trained network on the validation data did not improve after 150 epochs. Therefore, we evaluated the network after 150 epochs.

2.3 Experiments and Results

2.3.1 Experimental Setup

The performance of VParNet was compared with four state-of-the-art brain segmentation methods: FreeSurfer (version 6.0.0) [2, 48, 49], MALPEM [50], Joint label fusion (JLF) from the ANTs software package [52, 60], and RUDOLPH [61–63].

We also conducted an ablation analysis of the VParNet to see how different strategies affect the performance. The details of these network variations are reported in Table 2-II.

Table 2-II. Ablation analysis overview.

	Encoder	Normalization	Data augmentation	Combine multi-level feature maps
CNN-1	Residual	Instance		✓
CNN-2	Residual	Batch	✓	✓
CNN-3	Plain	Instance	✓	✓
CNN-4	Residual	Instance	✓	
VParNet	Residual	Instance	✓	✓

2.3.2 Evaluation Metrics

In our experiments, we used the Dice similarity coefficient (DSC) [76], 95% Hausdorff distance (HD) [77], and absolute volume difference (AVD) to evaluate the accuracy of

the ventricle parcellation results.

Dice similarity coefficient

The Dice similarity coefficient (DSC) is a volume-based metric broadly used to evaluate segmentation results. The DSC between two binary segmentation masks S and T is defined as

$$\text{DSC} = \frac{2|S \cap T|}{|S| + |T|}.$$

The DSC values are in the range $[0, 1]$, where 1 indicates perfect overlap between S and T and 0 indicates no overlap at all. We calculated the DSC between the automatic segmentations and the manual delineations to evaluate the performances of the different methods.

95% Hausdorff distance

Hausdorff distance (HD) is a distance-based metric used to measure the boundary distance between two segmentations. It is the longest of all the distances from a point in one point set to the closest point in the other point set. The HD between two sets of points A and B can be defined as

$$\text{HD} = \max\left\{\max_{a \in A} d(a, B), \max_{b \in B} d(b, A)\right\}, \tag{2.1}$$

where a and b are points from sets A and B, respectively, and $d(a, B)$ is the distance between a point a and the set B, which is defined as:

$$d(a, B) = \min_{b \in B} |a - b|. \tag{2.2}$$

Lower values of HD indicate higher segmentation accuracy. We used the 95% distance to calculate the HD, since HD is sensitive to outliers.

Absolute volume difference

The absolute volume difference (AVD) between the volume of the segmentation result V_S and the volume of the ground truth V_T is defined as

$$\text{AVD} = \frac{|V_S - V_T|}{V_T} \times 100\%.$$

Lower values of AVD indicate better volumetric agreement.

2.3.3 Comparison to Other Methods

The objective of the proposed method is to provide accurate parcellation of the ventricular system in both healthy controls and patients with mild to severe cases of ventriculomegaly. All the methods, including FreeSurfer (version 6.0.0), MALPEM, JLF, RUDOLPH, and VParNet, were run in an automatic manner.

Ventricle parcellation of healthy brains

In this experiment, we ran each method on the 30 testing subjects from the NMM data set, investigating their performances on healthy subjects. Although MALPEM and FreeSurfer have built-in skull-stripping methods, we provided our skull-stripped data to them for consistency with the processing; we note that in our experience, FreeSurfer produced better results with our provided skull-stripped data than the built-in skull-stripping. The 15 NMM images used to train the networks served as the atlases for MALPEM, JLF, and RUDOLPH. We used symmetric image normalization (SyN) [51] for non-rigid registration in JLF and RUDOLPH.

A visual comparison of the ventricle parcellation results produced by the five methods is shown in Figure 2-5. We only present the ventricle labels of the results from FreeSurfer, MALPEM, JLF, and RUDOLPH. We observe that MALPEM slightly over-segmented the
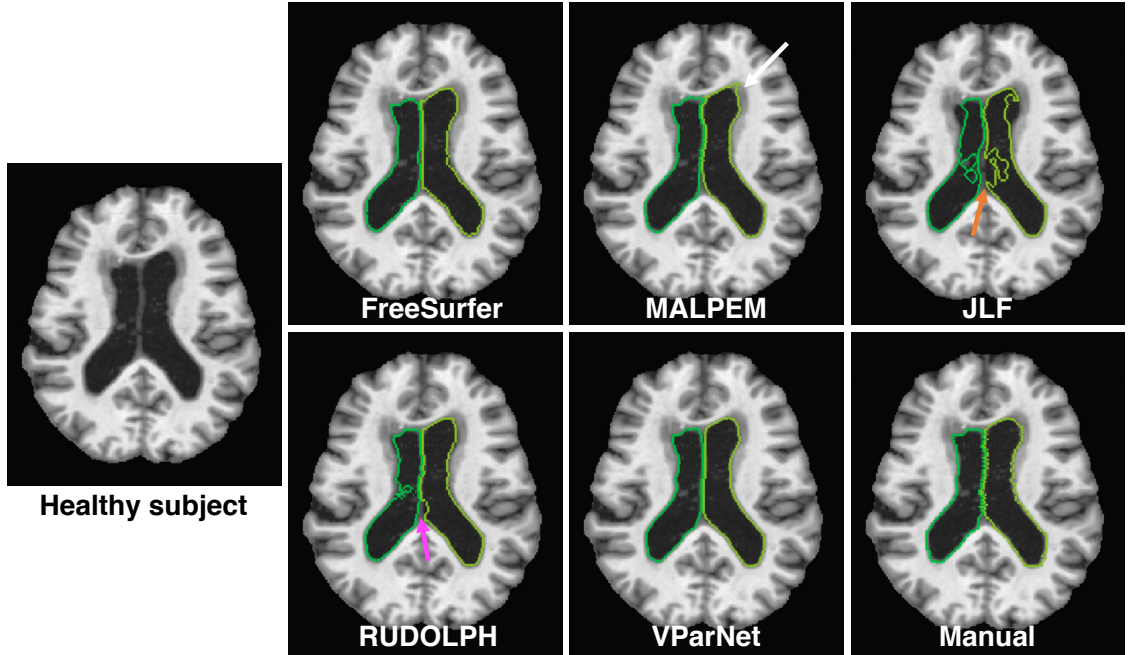
Figure 2-5. Visual comparison of the five segmentation methods for one healthy subject from the NMM data set. The arrows point to inaccurate segmentation regions.

left lateral ventricle (see the white arrow in Figure 2-5), and JLF and RUDOLPH did not capture the ventricle boundaries near the septum pellucidum on this subject (see the orange and magenta arrows in Figure 2-5). Boxplots of the DSC, 95% HD, and AVD over 30 testing MRIs from the NMM data set are shown in Figure 2-6, left side. Detailed mean and standard deviation values of each metric are reported in Table 2-III. We conducted a paired Wilcoxon signed-rank test [78] with an α-level of 0.005 and mark the significant difference between VParNet and each of the other methods in asterisks at the top/bottom of each plot. VParNet outperformed FreeSurfer in each evaluation metric except AVD of the third and fourth ventricles. Comparing to MALPEM, VParNet achieved competitive results in terms of DSC and AVD, and better results on the lateral ventricles in terms of 95% HD. Comparing to JLF and RUDOLPH, VParNet produced overall higher DSC and lower 95% HD and AVD.
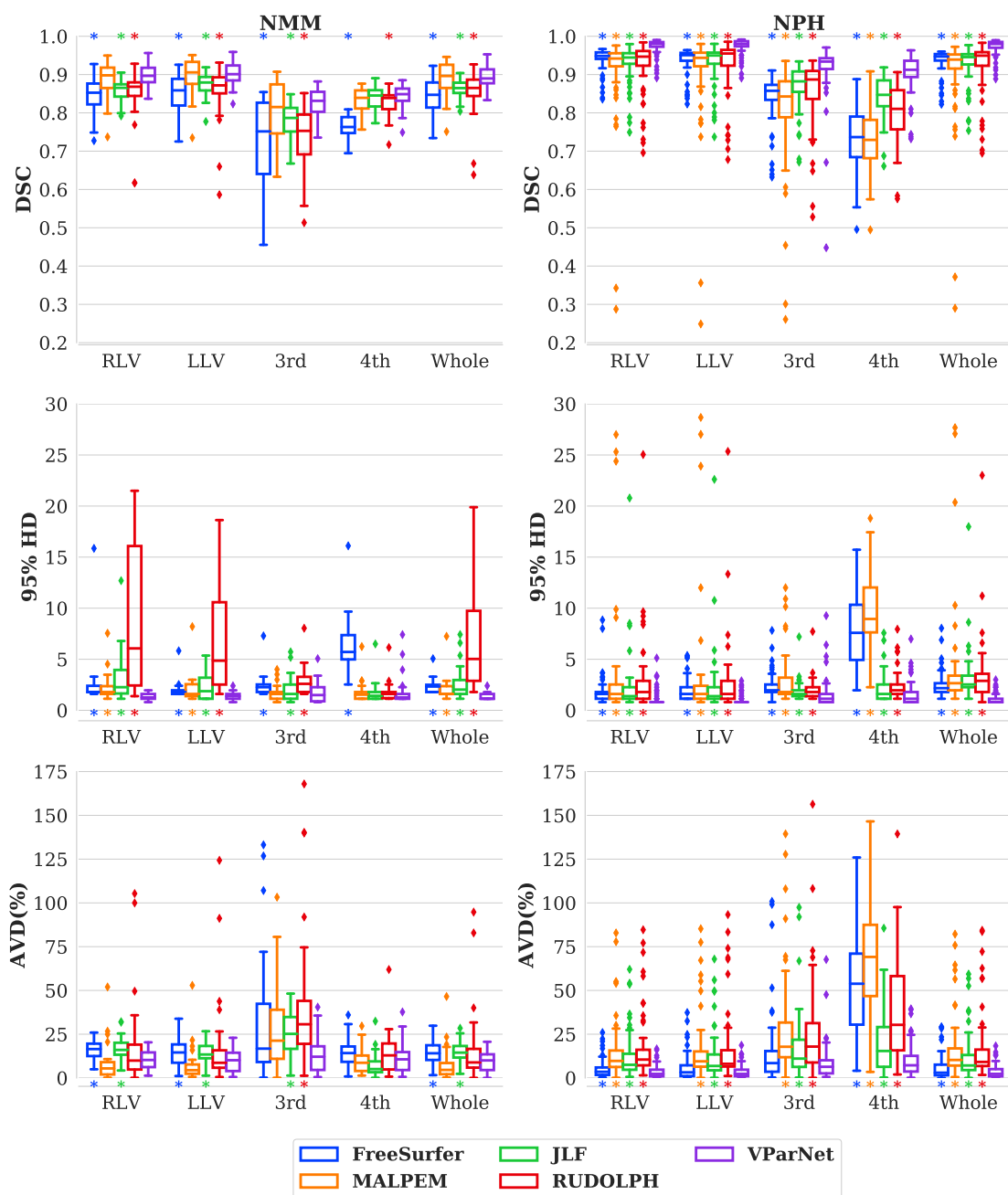
Figure 2-6. Boxplots of the Dice similarity coefficient (DSC), 95% Hausdorff distance (HD, in mm), and absolute volume difference (AVD) over 30 T1-w MRIs from the NMM data set (left side) and 65 T1-w MRIs from the NPH data set (right side). Ventricular system key: Right lateral ventricle (RLV), left lateral ventricle (LLV), third ventricle (3rd), fourth ventricle (4th), and whole ventricular system (Whole). The asterisks mean significantly different (p-value < 0.005) from VParNet results.

Table 2-III. The mean (standard deviation) of Dice similarity coefficient (DSC), 95% Hausdorff distance (HD, in mm), and absolute volume difference (AVD) over 30 T1-w MRIs from the NMM data set and 65 T1-w MRIs from the NPH data set generated by the five algorithms. Ventricular system key: Right lateral ventricle (RLV), left lateral ventricle (LLV), third ventricle (3rd), fourth ventricle (4th), and whole ventricular system (Whole). The asterisks mean the results of the corresponding method are significantly different (p-value < 0.005) from VParNet results.

	RLV	LLV	3rd	4th	Whole
NMM:					
DSC:					
FreeSurfer	0.847*(0.05)	0.852*(0.04)	0.729*(0.22)	0.763*(0.03)	0.845*(0.05)
MALPEM	0.889(0.05)	0.897(0.04)	0.803(0.08)	0.833(0.03)	0.891(0.04)
JLF	0.859*(0.03)	0.877*(0.03)	0.779*(0.04)	0.837(0.03)	0.865*(0.02)
RUDOLPH	0.851*(0.07)	0.857*(0.07)	0.735*(0.09)	0.827*(0.04)	0.853*(0.06)
VParNet	0.898(0.03)	0.900(0.03)	0.824(0.04)	0.840(0.03)	0.895(0.03)
95% HD:					
FreeSurfer	2.496*(2.6)	1.991*(0.77)	2.432*(1.0)	6.204*(2.6)	2.404*(0.68)
MALPEM	2.157*(1.3)	2.063*(1.3)	1.708(0.72)	1.728(1.0)	2.051*(1.1)
JLF	3.176*(2.5)	2.387*(1.3)	2.077*(1.2)	1.706(1.0)	2.555*(1.6)
RUDOLPH	8.741*(7.1)	7.263*(5.7)	2.811*(1.3)	1.799(0.95)	7.123*(5.8)
VParNet	1.350(0.28)	1.387(0.36)	1.736(1.0)	1.834(1.4)	1.366(0.32)
AVD(%):					
FreeSurfer	15.4*(5.5)	14.6*(7.2)	36.0(46)	13.9(8.1)	14.2*(6.8)
MALPEM	8.34(11)	7.46(9.9)	28.0(25)	9.52(7.0)	7.35(9.3)
JLF	16.9*(6.5)	14.1*(6.2)	25.1*(13)	7.04(6.6)	15.0*(6.0)
RUDOLPH	18.2(26)	17.7(27)	43.5*(42)	13.7(12)	16.3(22)
VParNet	10.1(5.2)	10.0(6.3)	13.7(11)	11.0(8.3)	9.72(5.6)
NPH:					
DSC:					
FreeSurfer	0.941*(0.03)	0.938*(0.03)	0.839*(0.06)	0.735*(0.08)	0.936*(0.03)
MALPEM	0.910*(0.12)	0.909*(0.12)	0.808*(0.13)	0.727*(0.08)	0.905*(0.11)
JLF	0.929*(0.05)	0.931*(0.05)	0.869*(0.06)	0.842*(0.05)	0.929*(0.05)
RUDOLPH	0.925*(0.07)	0.927*(0.07)	0.856*(0.09)	0.801*(0.08)	0.925*(0.07)
VParNet	0.974(0.02)	0.975(0.02)	0.917(0.07)	0.904(0.05)	0.973(0.02)
95% HD:					
FreeSurfer	1.810*(1.4)	1.874*(1.0)	2.383*(1.1)	7.034*(3.2)	2.418*(1.2)
MALPEM	3.067*(5.2)	3.000*(5.5)	2.784*(2.3)	9.654*(3.5)	3.969*(5.0)
JLF	2.156*(2.8)	2.136*(3.0)	1.794*(0.85)	2.140*(1.3)	3.121*(2.2)
RUDOLPH	2.761*(3.5)	2.570*(3.5)	1.870*(0.94)	2.346*(1.3)	3.335*(3.0)
VParNet	1.092(0.75)	1.030(0.49)	1.566(1.4)	1.560(1.1)	1.086(0.54)
AVD(%):					
FreeSurfer	5.37*(5.5)	5.78(7.4)	14.7(21)	54.5*(29)	5.98*(7.0)
MALPEM	15.2*(17)	15.4*(18)	27.1*(29)	71.0*(33)	16.1*(17)
JLF	11.9*(13)	11.9*(14)	16.8*(18)	20.7*(18)	11.5*(13)
RUDOLPH	16.4*(18)	16.2*(20)	26.9*(32)	39.7*(29)	16.4*(19)
VParNet	3.55(3.6)	3.47(3.7)	8.89(10)	10.1(9.3)	3.62(3.8)

Ventricle parcellation of NPH patients

In this experiment, we ran each method on the 65 testing subjects from the NPH data set. In FreeSurfer, we turned the "bigventricles" switch "on" to account for the enlarged ventricles. For the non-rigid registration method SyN in JLF and RUDOLPH, we set the step size of the gradient descent to 0.3 to enable larger deformations.

Examples of the ventricle parcellation results obtained from the five methods on mild, moderate, and severe NPH subjects are shown in Figure 2-7 and Figure 2-8. The subject in Figure 2-7(a) is a mild NPH case with relatively small ventricles. MALPEM, JLF, and RUDOLPH tended to over-segment the ventricles of this subject (see the white, orange, and magenta arrows in Figure 2-7(a)), while VParNet performed better on the boundaries. VParNet produced smoother segmentations than FreeSurfer. Figure 2-7(b) shows a moderate case of NPH, where most methods performed well on ventricle parcellation. The yellow and white arrows point to inaccurate segmentations on the boundaries from FreeSurfer and MALPEM, respectively. Figure 2-8 show two NPH subjects with severely enlarged ventricles. FreeSurfer, MALPEM, and JLF failed to identify the boundaries of the enlarged ventricles (see the yellow, white, and orange arrows), and RUDOLPH slightly underestimated the size of the left lateral ventricle in one subject (see the magenta arrow in Figure 2-8(a)).

Boxplots of DSC, 95% HD, and AVD are presented in Figure 2-6, right side. Detailed mean and standard deviation values of each metric are reported in Table 2-III. The proposed VParNet performed better than all the other methods on each ventricle label in terms of DSC and 95% HD. These results were statistically significant in a paired Wilcoxon signed-rank test with an α-level of 0.005. Regarding AVD, VParNet significantly outperformed the other methods on each label, except on the left lateral ventricle and the third ventricle results generated by FreeSurfer, where the difference did not reach statistical significance. In summary, these boxplots demonstrate that VParNet produced
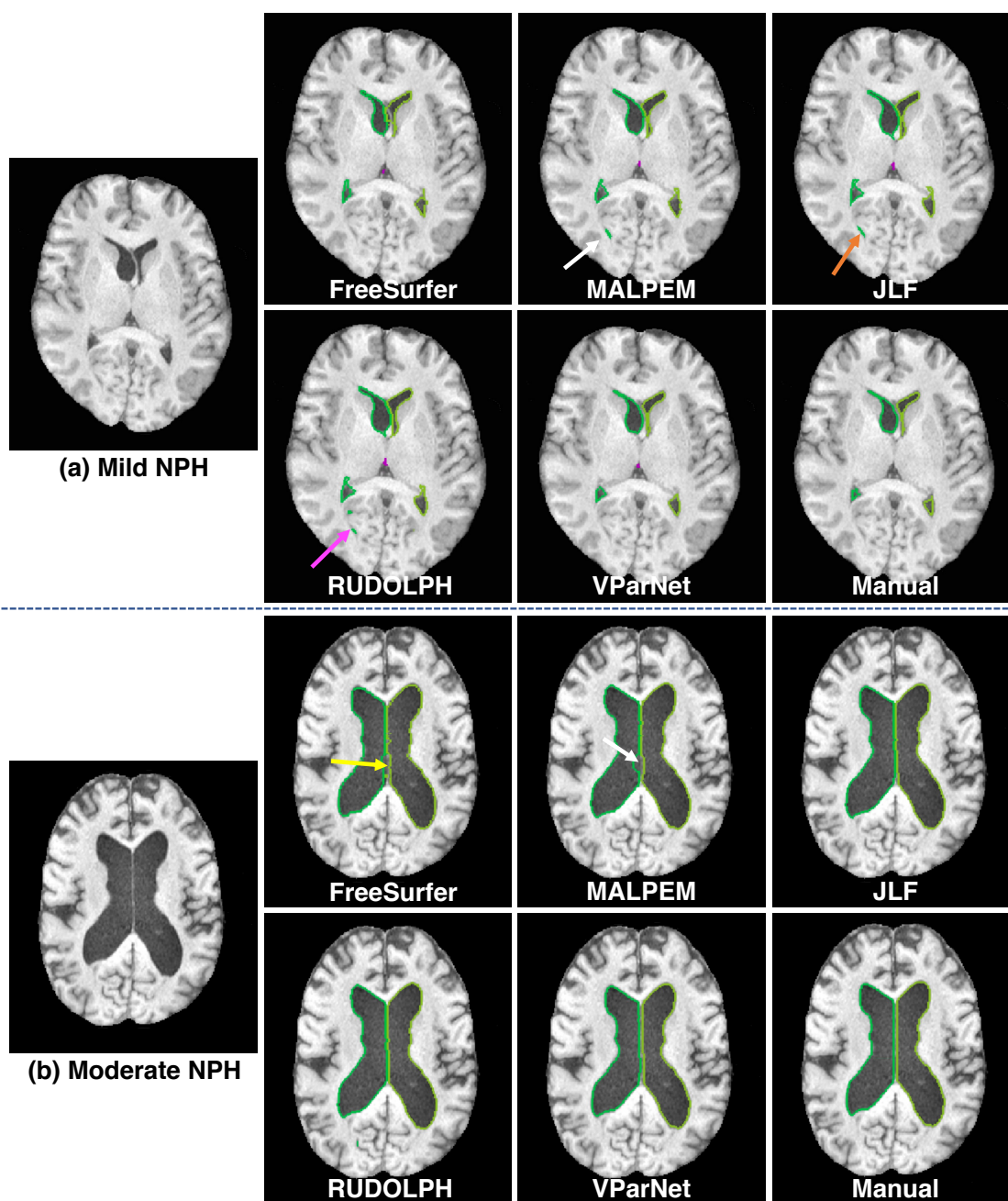
Figure 2-7. Visual comparison of the five segmentation methods for one mild and one moderate NPH subjects. The arrows point to inaccurate segmentation regions.
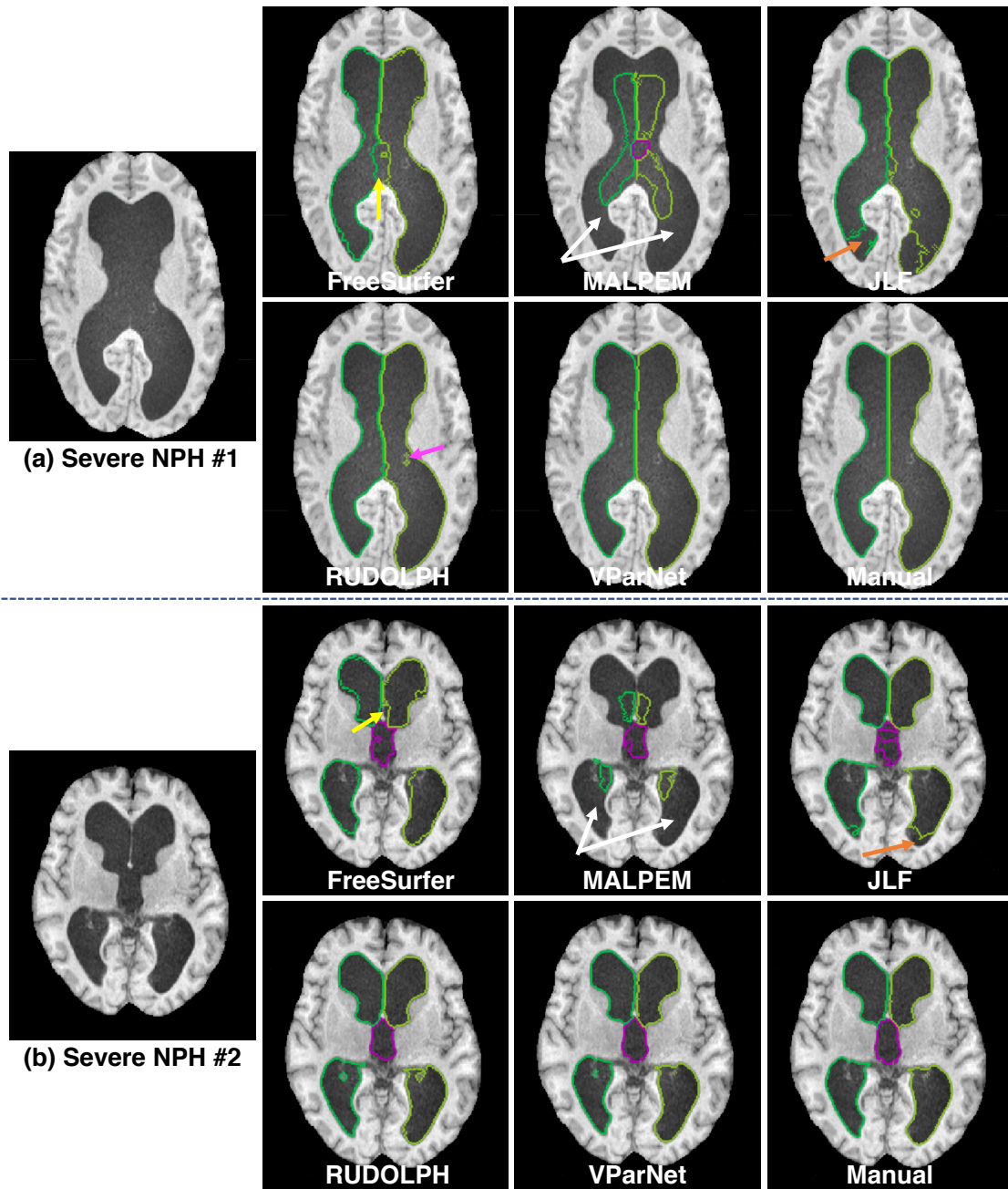
Figure 2-8. Visual comparison of the five segmentation methods for two severe NPH subjects. The arrows point to inaccurate segmentation regions.

more accurate ventricle parcellation results on subjects with highly variable ventricle sizes, from healthy to severe ventriculomegaly.

2.3.4 Ablation Study

We conducted an ablation analysis of the VParNet to see how different strategies affect the performance of the network (see Table 2-II for the details of the network variations). Table 2-IV shows the results of the ablation analysis. The training set was the same in each network, except that the training images were white matter peak normalized before use in CNN-2, which used batch normalization instead of instance normalization.

When comparing CNN-1 and VParNet, we observe that adding data augmentation improves the parcellation performance in terms of both DSC and 95% HD. The improvements show significance in a paired Wilcoxon signed-rank test with an α-level of 0.005. The comparison between CNN-2 and VParNet demonstrates that using instance normalization produces comparable parcellation results with the network using batch normalization. The performance of the instance normalization comes from the training and testing not requiring intensity normalization. Comparing to CNN-3, VParNet has broadly higher DSC and lower 95% HD and AVD. The results show significant improvements on the AVD of the lateral ventricles, indicating the contribution of the short skip connection within the encoder block. Comparing CNN-4 and VParNet, both networks converge to similar validation loss. However, VParNet used only 5 epochs to reduce the validation loss to 0.1, while CNN-4 used 24 epochs. This shows that combining multi-level feature maps speeds up network convergence.

2.3.5 Five-fold cross validation

We applied 5-fold cross validation on our data sets to evaluate the proposed network. The 50 MRIs from the NMM data set and the 95 MRIs from the NPH data set were

Table 2-IV. The mean (standard deviation) of Dice similarity coefficient (DSC), 95% Hausdorff distance (HD, in mm), and absolute volume difference (AVD) over 95 testing images (30 from NMM and 65 from NPH) in the ablation study. Ventricular system key: Right lateral ventricle (RLV), left lateral ventricle (LLV), third ventricle (3rd), fourth ventricle (4th), and whole ventricular system (Whole). The asterisks mean significantly different (p-value < 0.005) from VParNet results.

	RLV	LLV	3rd	4th	Whole
DSC:					
CNN-1	0.944*(0.05)	0.947*(0.04)	0.880*(0.06)	0.875*(0.05)	0.944*(0.04)
CNN-2	0.946*(0.05)	0.947*(0.05)	0.887(0.07)	0.875(0.06)	0.944*(0.05)
CNN-3	0.948*(0.04)	0.951(0.04)	0.891(0.07)	0.881(0.05)	0.947(0.04)
CNN-4	0.949(0.04)	0.951(0.04)	0.887(0.07)	0.882(0.06)	0.948(0.04)
VParNet	0.950(0.04)	0.951(0.04)	0.887(0.08)	0.884(0.05)	0.948(0.04)
95% HD:					
CNN-1	1.738*(1.7)	1.573*(1.1)	1.727*(1.0)	1.958(1.5)	1.454*(1.2)
CNN-2	1.367*(0.99)	1.281(0.89)	1.641(1.1)	1.663(1.3)	1.311*(0.80)
CNN-3	1.232(0.79)	1.158(0.53)	1.489(1.0)	2.730(11)	1.238(0.62)
CNN-4	1.393(2.2)	1.143(0.48)	1.990*(2.4)	1.689(1.3)	1.162(0.49)
VParNet	1.173(0.65)	1.143(0.48)	1.620(1.3)	1.648(1.2)	1.174(0.50)
AVD(%):					
CNN-1	5.54(5.2)	5.26(5.6)	12.3*(11)	11.0(9.8)	5.37(5.1)
CNN-2	6.51(6.7)	6.68*(6.8)	10.1(10)	13.1(11)	6.66*(6.6)
CNN-3	6.02*(5.6)	5.80*(5.6)	9.85(9.1)	11.5(9.0)	5.90*(5.5)
CNN-4	5.76(5.1)	5.45*(5.4)	10.5(9.9)	11.1(9.1)	5.57(5.1)
VParNet	5.62(5.2)	5.54(5.5)	10.4(11)	10.4(9.0)	5.55(5.2)

split into 5 groups. Each group contained 10 NMM images and 19 NPH images. Firstly, we sorted the 50 NMM images and 95 NPH images separately by the volume of the ventricular system. Secondly, the sorted NMM and NPH data sets were evenly divided into 10 and 19 small groups, respectively. Each small group contained 5 MRIs with similar ventricle sizes. Thirdly, for each small group, we randomly permutated the 5 MRIs and assigned them to our final 5 groups (one image per group). Therefore, each of the 5 cross-validation groups will cover a wide spectrum of ventricle sizes from both data sets.

During cross validation, we trained VParNet 5 times using different training data. The network trained in the ith time, which we referred to as TestFold-i, took the ith group

as testing set and the remaining 4 groups as training set. Among the training set, 5 NMM images and 10 NPH images were used for validation. Quantitative evaluation of the 5 networks is shown in Table 2-V. We conducted a Mann–Whitney U test [79] with α-level of 0.01 to compare each pair of the 5 networks. The differences between any two networks in terms of the three evaluation metrics are not significant except the AVD of the right lateral ventricle between TestFold-3 and TestFold-4, which has $p = 0.0051$.

In this experiment, we evaluated the performance of the proposed network trained on different partitions of the data. The 5 networks produced similar parcellation results in terms of DSC, 95% HD, and AVD, indicating that the proposed network performs well on unseen data.

Table 2-V. The mean (standard deviation) of Dice similarity coefficient (DSC), 95% Hausdorff distance (HD, in mm), and absolute volume difference (AVD) for each cross-validation group over 10 NMM data and 19 NPH data. TestFold-i means the ith group was taken as testing data and the network was trained on the remaining 4 groups. Ventricular system key: Right lateral ventricle (RLV), left lateral ventricle (LLV), third ventricle (3rd), fourth ventricle (4th), and whole ventricular system (Whole).

	RLV	LLV	3rd	4th	Whole
DSC:					
TestFold-1	0.949(0.04)	0.951(0.04)	0.892(0.07)	0.889(0.05)	0.949(0.04)
TestFold-2	0.952(0.04)	0.951(0.04)	0.883(0.09)	0.889(0.05)	0.949(0.04)
TestFold-3	0.948(0.05)	0.949(0.05)	0.888(0.09)	0.881(0.06)	0.946(0.05)
TestFold-4	0.943(0.05)	0.948(0.04)	0.865(0.09)	0.890(0.04)	0.944(0.04)
TestFold-5	0.952(0.04)	0.953(0.04)	0.899(0.05)	0.897(0.04)	0.950(0.04)
95% HD:					
TestFold-1	1.013(0.24)	0.989(0.29)	1.397(0.70)	1.657(1.27)	1.211(0.84)
TestFold-2	1.075(0.42)	1.059(0.38)	1.312(0.64)	1.463(1.09)	1.097(0.44)
TestFold-3	1.087(0.60)	1.085(0.42)	1.288(0.81)	1.490(0.85)	1.111(0.40)
TestFold-4	1.734(2.57)	1.065(0.30)	1.868(1.59)	1.181(0.32)	1.301(0.61)
TestFold-5	1.071(0.45)	1.049(0.37)	1.317(0.79)	1.512(0.78)	1.050(0.30)
AVD(%):					
TestFold-1	5.12(4.4)	5.05(4.0)	9.11(7.9)	9.13(8.7)	4.69(3.4)
TestFold-2	4.31(4.1)	4.42(5.0)	11.4(12)	8.54(9.1)	4.36(4.7)
TestFold-3	3.97(5.8)	4.54(6.0)	9.83(11)	11.8(11)	4.11(5.7)
TestFold-4	6.35(5.2)	5.99(5.1)	12.8(12)	9.67(8.2)	6.10(5.0)
TestFold-5	4.60(4.1)	4.22(4.1)	8.57(7.5)	9.38(6.0)	4.27(3.5)

2.3.6 Pilot Study

We conducted a pilot study to apply VParNet to a clinical data set. The subjects in the data set were diagnosed with NPH and were treated by shunt surgery. This data set contains MR images before and after the treatment for each subject. After the shunt surgery, the patients have an implanted valve in a posterior-superior position under the skull, where signal dropout during MR image acquisition is expected (see the left-bottom figure of each axial slice in Figure 2-9). A total of 12 T1-w MR images from 6 subjects, including pre- and post-surgery data, were processed by VParNet trained from the NMM and NPH data sets. Two axial slices of one subject and the corresponding parcellation results are shown in Figure 2-9. In each axial slice, the top and bottom rows show the image and the parcellations before and after the shunt surgery, respectively. Visually, we did not find major failures for the parcellation results for all the images, although the network was not trained on data with implanted valves. However, we do observe some small regions in the area with the implanted valve that were mislabeled as lateral ventricles by VParNet (see yellow contour in Figure 2-9). Degraded deep network performance is expected when the testing data show different properties from the training data. To improve the accuracy of the parcellation in this data set, we could add a post-processing step to the algorithm to remove small false positive segmentation, or add new training data with implant valve to the network to improve its generalization.

2.4 Discussion

We have performed a comprehensive evaluation of our proposed VParNet using data from two different data sets, NMM and NPH. The accuracy of the ventricle parcellation was evaluated using DSC, 95% HD, and AVD between the automated parcellation and manual delineations. VParNet was trained on 15 healthy controls and 25 subjects with
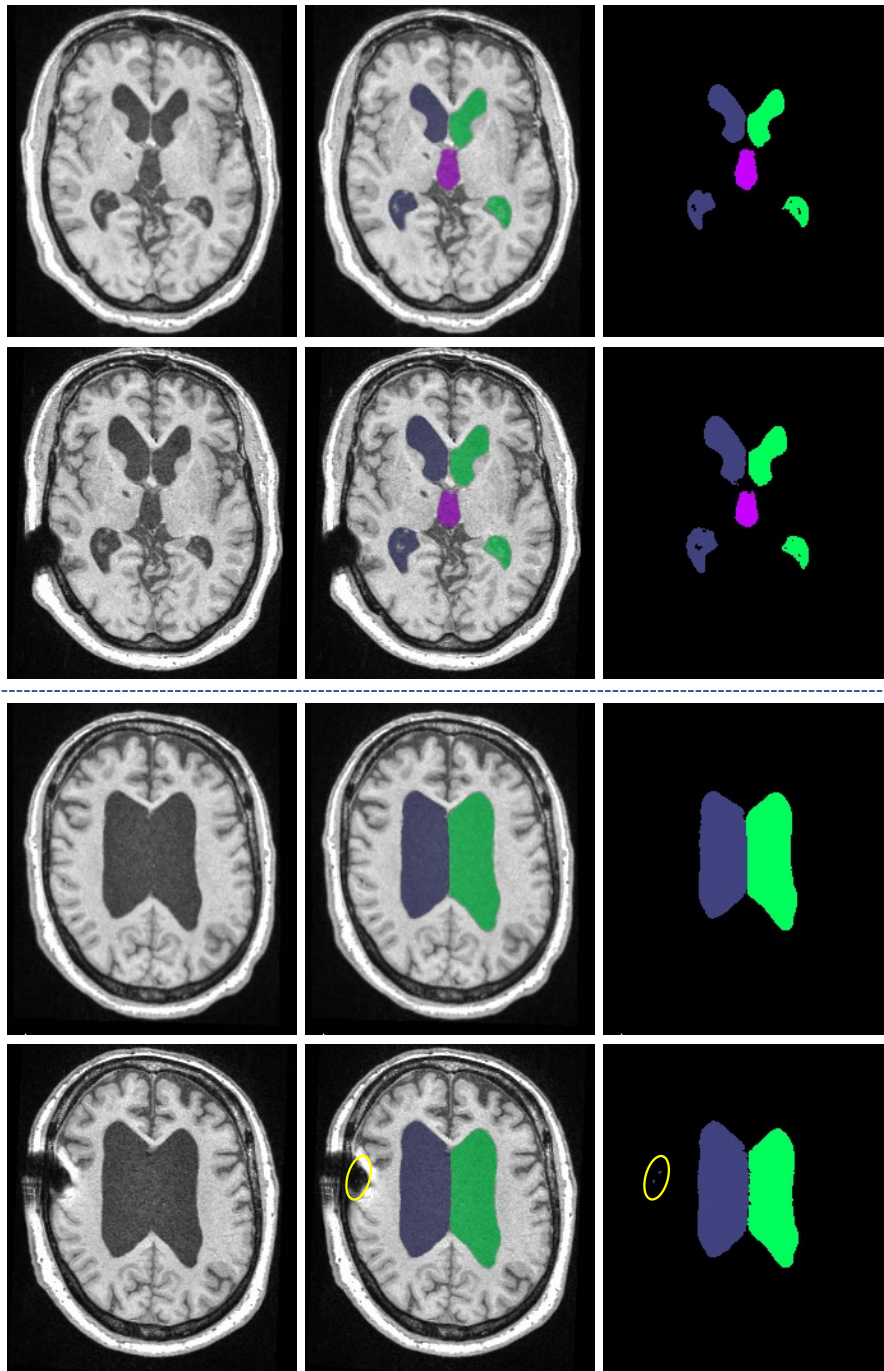
Figure 2-9. Two axial T1-w MR image slices and their corresponding VParNet parcellation results of a subject in a clinical study. In each slice, the top and bottom rows show the images before and after the shunt surgery, respectively. From left to right of each row: T1-w MR image, T1-w image overlaid by the parcellation result, and the parcellation result. The yellow contour points to small false positive regions.

NPH. When testing on the NMM data set, our method achieved competitive parcellation results compared to state-of-the-art brain segmentation algorithms, as shown in Figure 2-6, left side. The NPH data set was then used to demonstrate the robustness of VParNet to subjects with highly variable ventricle sizes and shapes. Our method produced significantly better results than the other methods in terms of DSC, 95% HD and AVD, as shown in Figure 2-6, right side. As shown in the boxplots, our method achieved better evaluation scores and lower standard deviations, demonstrating the consistency of the network performance on healthy and NPH subjects.

Considering the performance of VParNet on the two data sets, the network obtained overall higher DSC and lower AVD in the NPH data set compared to the healthy control data set. The reason is that these two metrics are biased by the size of the segmented structure. In the NMM data set, the ventricular volume is smaller, ranging from 12 ml to 132 ml with mean value being 51 ml, while the ventricular volume in the NPH data set ranges from 10 ml to 400 ml with mean value being 120 ml. A similar bias is seen when considering the size of different ventricular compartments; VParNet reached a better evaluation score on the lateral ventricles than the third and the fourth ventricles in terms of DSC and AVD.

Automated parcellation of the ventricular system in MRIs of patients with enlarged ventricles is challenging due to the variations of ventricle volumes and shapes across individual subjects. Many approaches have been proposed to segment the human brain in MRI, and some of them are capable of parcellating the ventricular system into its four main cavities (see Table 2-I). Conventional atlas-based methods require deformable image registration, which is typically time-consuming. Some recently developed deep learning-based image registration algorithms [80–82] can speed up the registration process. However, the enlarged ventricles can cause significant registration errors, as the difference between the atlas and the subject image is big. FreeSurfer has a special "bigventricles" option to account for enlarged ventricles in hydrocephalus and late stage

40

Alzheimer's cases. It is worth noting that we used this special option when running FreeSurfer on the NPH data set. We also experimented with running FreeSurfer in its default setting, and on some subjects with severely enlarged ventricles the program failed to output segmentation results at all. For this FreeSurfer experiment, the mean DSC of the whole ventricular system on the remaining testing images was 0.77, while in FreeSurfer with "bigventricles" flag, the mean DSC was 0.94.

In our experiments, we also processed our data with BrainSuite [58], MUSE [57], and NLSS [59]. BrainSuite produced segmentation results with mean DSC of 0.6 in the NPH data set, which did not compare favorably with the other methods. MUSE can be run remotely on a web platform; however, we were not permitted to upload our NPH data due to health care privacy considerations. The multi-atlas based method NLSS takes about 36 hours to process one subject, which was not efficient and made it challenging for us to complete the entire test data set. Therefore, we did not report comparisons of these methods to the proposed network here; however, we note that none of these methods performed better than our network.

When exploring the variations of the network, the accuracy of VParNet results was statistically higher than the network without data augmentation. This demonstrates the value of data augmentation in tasks with limited amounts of training data. VParNet and the network without combining multi-level feature maps converged to a similar validation loss. However, VParNet converged much faster, indicating the importance of incorporating features at different levels. We also trained a network with batch normalization using the training data without white matter peak normalization; however, this network showed poor generalization, demonstrating the strength of instance normalization that does not require intensity normalization of the input images.

The typical processing time of the five methods is reported in Table 2-VI. All the programs were run on a 16-core 2.4 GHz CPU. A CPU version of VParNet required about two minutes to process one image, which is much lower than the several hours

41

Table 2-VI. Processing time comparison of different methods.

Method	Runtime per image
FreeSurfer	9.5 h
MALPEM	2 h
JLF	15 h
RUDOLPH	15 h
VParNet	2 min

required by the other methods. We note that this processing time can be brought down to approximately 30 seconds when run on an NVIDIA Tesla K80 GPU.

2.5   Summary

In this chapter, we presented a 3D U-Net for ventricle parcellation of brains with enlarged ventricles from MRI. We incorporated the deep residual learning technique in the encoder block to ease the degradation problems in training. The refinement of the classification that combines feature maps from multiple resolution levels can speed up network convergence. We used instance normalization instead of batch normalization to make the network invariant to linear transformation of the image intensities. Thus the training and testing images can be sent to the network without intensity normalization.

We showed that our proposed network achieved state-of-the-art performance on a healthy data set and significantly better performance than top competing methods on an NPH data set. The evaluation indicated the robustness of our network to high variability in ventricle shapes and sizes; from healthy to severe ventriculomegaly. Our method provides a tool for automated parcellation of the ventricular system in healthy subjects and patients with ventriculomegaly, enabling more sophisticated analyses of the ventricular structures in both health and disease.

Chapter 3

Brain thalamus parcellation using multimodal MRI

3.1 Introduction

3.1.1 Anatomy of Thalamus and Thalamic Nuclei

The thalamus is a key gray matter subcortical structure that is located between the cerebral cortex and mid-brain. It consists of two symmetric structures, one in each hemisphere. The shape of each half of the thalamus is like a walnut and the length is about 30 mm. The thalamus is the main relay station for the brain and controls how sensory and motor signals are passed through different parts of the cerebral cortex [83]. It is a highly complex structure and can be subdivided into numerous thalamic nuclei that different in both histology and function. Figure 3-1 shows an example of a manually delineated thalamic nuclei mask overlaid on a T1-w MPRAGE MR image and its 3D rendering visualization.

The subdivision of the thalamus into nuclei depends on histological and functional criteria. The thalamus parcellation and delineation schemes are very different when it

comes to detailed interpretations [84]. For example, the software package FreeSurfer [2] used a probabilistic atlas [85] built using ex-vivo brain MRI scans and histological data. The atlas parcellate the thalamus into 26 different nuclei. The Morel atlas of the thalamus [86, 87] identified the thalamic nuclei based on multiarchitectonic parcellation using post-mortem human brain data. The finest level includes around 30 thalamic nuclei labels. However, since the size of thalamus is small and many of these nuclei cannot be reliably identified from medical imaging. In this thesis, we group some nuclei together and focus on the following 6 thalamic nuclei: anterior nucleus (ANT), medial dorsal nucleus (MDN), ventral nuclear group (VNG), pulvinar nucleus (PUL), medial geniculate nucleus (MGN), and lateral geniculate nucleus (LGN). These nuclei are more readily distinguish by multimodal MRI.

Each nucleus in the thalamus connects to one or more parts of the cerebral cortex, and therefore is associated with different neurological functions. For example, the anterior nucleus is connected to cingulate gyrus which is involved in processing emotions and behavior regulation [88]. The medial dorsal nucleus shows connections to the prefrontal cortex [89]. The pulvinar nucleus connects to parietal, occipital, and temporal lobes [90]. The lateral geniculate shows strong connections to primary visual cortex in occipital lobe [91]. Since the thalamus acts as a relay station that connects to different brain structures, damage of the thalamus is related to many brain disorders. Volume changes in thalamic nuclei are of great interest in many neurodegenerative studies, such as Alzheimer's disease [25], multiple sclerosis [26], and traumatic brain injury [27]. Thus, accurate and consistent segmentation of the thalamus would be beneficial to better understand and manage the progression and effect of neurological diseases.
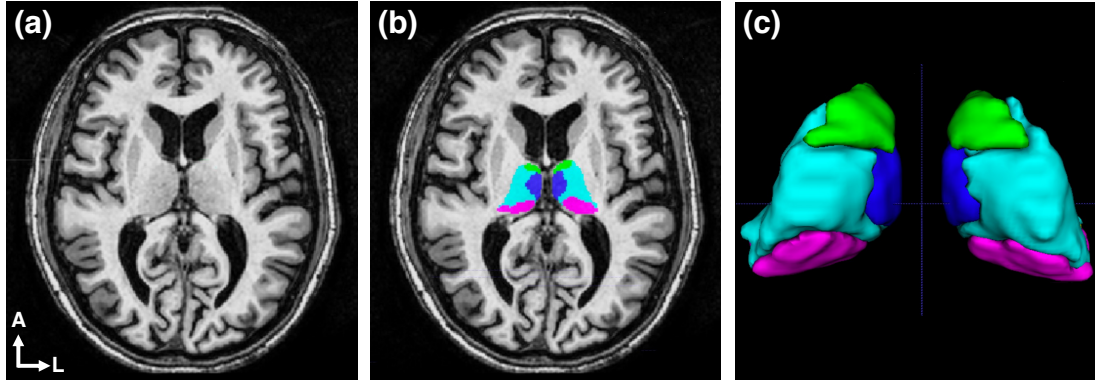
Figure 3-1. Examples of thalamus anatomy. (a) An axial slice of a T1-w MRI. (b) Thalamic nuclei mask overlaid on (a). (c) 3D rendering of the thalamic nuclei. The color of the labels correspond to thalamic nuclei as follows: anterior nucleus (green), medial dorsal nucleus (blue), ventral nuclear group (cyan), and pulvinar nucleus (magenta). A = anterior; L = left.

3.1.2   Diffusion Weighted Imaging

Automatic parcellation of the thalamus into different nuclei using MRI has been broadly explored [92–95]. However, standard structure MRI, such as T1-w MR images, has poor contrast within the thalamus to distinguish the thalamic nuclei [94, 95]. Many thalamus parcellation algorithms [92, 93, 96, 97] have reported the usage of information derived from diffusion weighted imaging (DWI). DWI is an MRI technique that is able to characterize the diffusion properties of water in the underlying tissue [98, 99]. In ordered structures such as nervous tissues or muscle tissues, water tends to diffuse along a particular direction and results in anisotropic diffusion. The DWI signal is based on measuring the movement of water molecules within a voxel of tissue by applying a preset gradient pulse, which determines a gradient direction.

Diffusion tensor imaging (DTI) [100, 101] is a special DWI technique that models the diffusion anisotropy in each voxel as a diffusion tensor. A diffusion tensor, denoted by D, is a 3×3 matrix that is symmetric and positive definite because of the nature of the diffusion process. The relationship between the DWI signal and the diffusion tensor is

given by the following equation:

$$S(\boldsymbol{g}, b) = S_0 \exp\left(-b\boldsymbol{g}^T \boldsymbol{D}\boldsymbol{g}\right) \tag{3.1}$$

where $\boldsymbol{g}$ is the unit gradient direction, b is the b-value, which is a key parameter in DWI that measures the degree of diffusion weighting and has a unit of s/mm². S_0 is the baseline signal without any diffusion weighting and $\boldsymbol{D}$ is the diffusion tensor. Since the diffusion tensor is a 3×3 symmetric matrix, it has 6 unknown coefficients to be estimated. Therefore, the DTI requires at least 6 DW images and one b0 image ($b = 0$) to solve $\boldsymbol{D}$ in Equation 3.1. Typically, a set of 6–32 gradient directions are used for DTI reconstruction [102]. The diffusion tensor can be estimated by least-squares methods [103].

From the estimated diffusion tensor, we can perform eigendecomposition to diagonalize the tensor and obtain three eigenvalues $\lambda_i, i \in \{1, 2, 3\}$, with $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and three corresponding eigenvectors $\boldsymbol{e}_i, i \in \{1, 2, 3\}$. The principal eigenvector $\boldsymbol{e}_1$ can be visualized as the direction of the largest diffusion. Several quantities can be calculated from the eigenvalues of $\boldsymbol{D}$ to measure the diffusion (See Figure 3-2 for examples of DTI features). For example, mean diffusivity is a measure of the degree of diffusion and is computed as

$$\text{MD} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{3.2}$$

The fractional anisotropy (FA) is a scalar value normalized between 0 and 1 that measures the degree of the anisotropy of a diffusion process. It is computed as

$$\text{FA} = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{3.3}$$

High FA values indicate the underlying tissue has highly organized fiber bundles along a
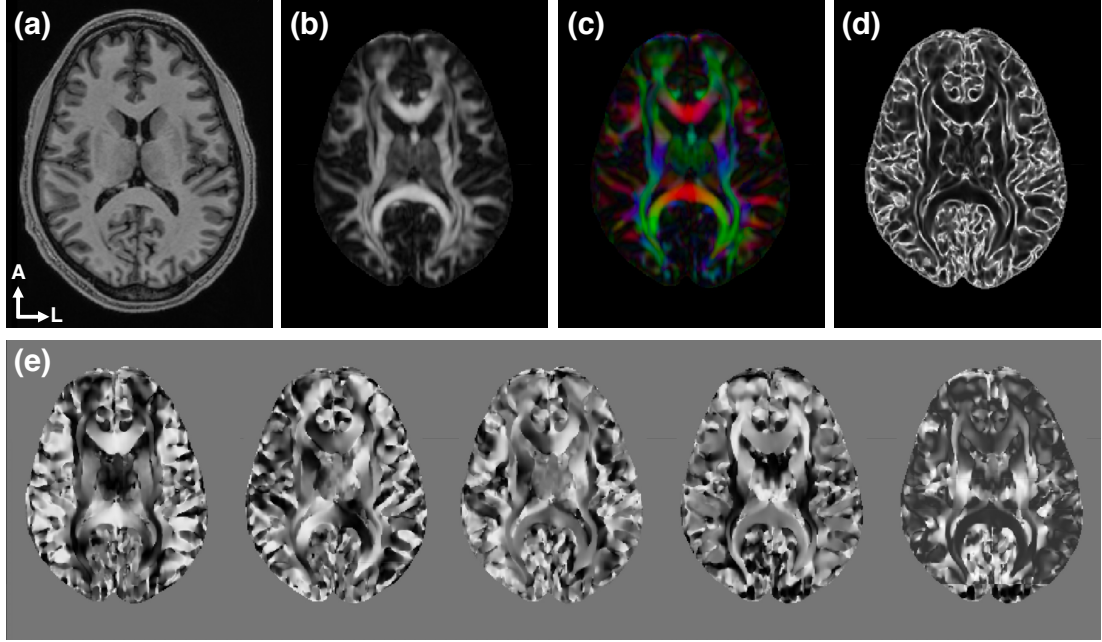
Figure 3-2. T1-w MR image and DTI feature images. (a) An axial slice of a T1-w MRI. (b) Fractional anisotropy (FA) image. (c) FA colored by primary eigenvector. (d) Knutsson edge map. (e) Knutsson 5D image.

single orientation. Low FA values indicate the underlying tissue has isotropic diffusion. Figure 3-2(b) shows an example of the FA image.

The diffusion tensor can be visualized as a colored FA map (see Figure 3-2(c)). The color in each voxel is conventionally encoded by the principal eigenvector e_1, where red represents left-right, green represents anterior-posterior, and blue represents inferior-superior. The brightness in this map is encoded by the FA map. The principal eigenvector e_1 is a useful feature because it represents the direction with largest diffusion. However, e_1 and its opposite direction $-e_1$ are considered as the same in DTI reconstruction. To avoid ambiguity, we can transform the 3D vector e_1 to a 5D vector k in the Knutsson space [104] using the following equation:

$$\boldsymbol{k}(\boldsymbol{e}_1) = (k_1, k_2, k_3, k_4, k_5) = (e_1^2 - e_2^2, 2e_1e_2, 2e_1e_3, 2e_2e_3, \frac{1}{\sqrt{3}}(2e_3^2 - e_1^2 - e_2^2)) \quad (3.4)$$

where $e_1 = (e_1, e_2, e_3)$. It can be seen that both e_1 and $-e_1$ are mapped to the same 5D vector in the Knutsson space. Figure 3-2(e) shows example of a Knutsson 5D map. The Knutsson mapping is continuous and preserves distances between vectors. From the Knutsson 5D map, we can compute a Knutsson edge map [105] which highlights the changes of the diffusion direction from one thalamic nuclei to another. The first step is to calculate the gradient matrix of k on the 3D spatial grid by

$$
G = \begin{bmatrix} \frac{\partial k_1}{\partial x} & \frac{\partial k_1}{\partial y} & \frac{\partial k_1}{\partial z} \\ \frac{\partial k_2}{\partial x} & \frac{\partial k_2}{\partial y} & \frac{\partial k_2}{\partial z} \\ \vdots & \vdots & \vdots \\ \frac{\partial k_5}{\partial x} & \frac{\partial k_5}{\partial y} & \frac{\partial k_5}{\partial z} \end{bmatrix}
\tag{3.5}
$$

Next, the edge map is characterized by the Frobenius norm of the gradient matrix G, as follows:

$$
||G||_F = \sqrt{\sum_i \sum_j G_{ij}^2}
\tag{3.6}
$$

An example of the Knutsson edge map is shown in Figure 3-2(d).

3.1.3 Related Work

Previous studies have developed algorithms to segment the thalamus and parcellate the thalamus into nuclei from MRI. Some methods rely only on standard structure MRIs. For example, Deoni et at. [106] proposed a modified k-means clustering algorithm to segment individual thalamic nuclei using T1-w and T2-w MRI. FreeSurfer [2, 85] used a probabilistic atlas built with histological data. The parcellation of the thalamus is done by solving a Bayesian inference problem from T1-w MRI. Su et al. [95] presented a multi-atlas segmentation technique based on standard or white-matter-nulled MPRAGE imaging to segment the thalamus into 12 thalamic nuclei. This automatic algorithm

uses a template that was created by registering and averaging 20 prior images as an intermediate space. The final segmentation is obtained by deformable registration between the template and the target volume, which dramatically reduces the registration time. In a recent work, Majdi et al. [107] presented a cascaded deep network that was modified from residual U-Net to segment thalamic nuclei from white-matter-nulled MPRAGE data.

Use of T1-w and T2-w MRIs alone provides poor contrast inside the thalamus to distinguish the nuclei. Therefore, many thalamic segmentation algorithms utilized features based on DTI. Ziyan et al. [96] presented a modified spectral clustering algorithm to identify the subdivision of the thalamus from DTI, where Markovian relaxation is used to handle spatial information. Duan et al. [108] proposed a semi-automatic algorithm for thalamus and thalamic nuclei segmentation from DTI. They used a mean shift clustering algorithm that supports hierarchical clustering. Glaister et al. [93] presented a hierarchical random forest for automatic thalamus parcellation. The algorithm used T1-w and T2-w MRIs, features derived from DTI, and a thalamic nuclei prior. Battistella et al. [94] presented a thalamic segmentation framework based on local diffusion properties and spatial features derived from diffusion MRI. The spherical harmonic decomposition based orientation distribution function was incorporated to provide better differentiation of the complex intra-thalamic microstructure.

3.1.4 MR Image Harmonization

In most brain structure segmentation algorithms based on deep learning, some MR images with corresponding ground truth labels, referred to as training data, are used to train a model and then the model can be applied to testing data. However, the contrast variation between the training and testing data can degrade the performance of the trained model on the testing data [20], which can be formalized as a domain shift

problem [109].

To address the domain shift problem, MR image harmonization, an image–to–image translation technique, can be used to transform the intensity from one domain to another and therefore can match the contrast between the training and testing images. CALAMITI [28, 110, 111], a deep learning based image harmonization algorithm, learns a disentangled latent space containing both anatomical and contrast information. To harmonize the MR image to a different domain, one can combine the anatomical information from the input image and the contrast information in the target domain to generate a new image. Figure 3-3 shows an example of CALAMITI result. Figure 3-3(a) and (b) are the source and target T1-w MR images. The goal is to manipulate the intensity of the source image and match the contrast between the source and the target. After applying CALAMITI, the harmonized image (Figure 3-3(c)) looks similar to Figure 3-3 in terms of contrast. The contrast latent space in CALAMITI is represented by a 2D vector, which can be directly visualized as a point in the standard 2D Cartesian coordinate system. The bottom figure in Figure 3-3 illustrates the contrast latent space before and after CALAMITI. The point after harmonization (blue circle) is closer to the target point (blue square) compared to the original point (orange circle).

CALAMITI can be used to transform the image contrast between different domains. It is worth noting that while the harmonization algorithm manipulates image intensity, the underlying brain anatomy should be maintained. In this chapter, we investigate the impact of CALAMITI on the thalamus parcellation task.

3.1.5 Contributions

In this chapter, we present a 3D U-Net method to segment the thalamus and parcellate it into different thalamic nuclei from multimodal MRI [24]. The input to the network includes T1-w MR image and features derived from DTI. We explore the impact of an
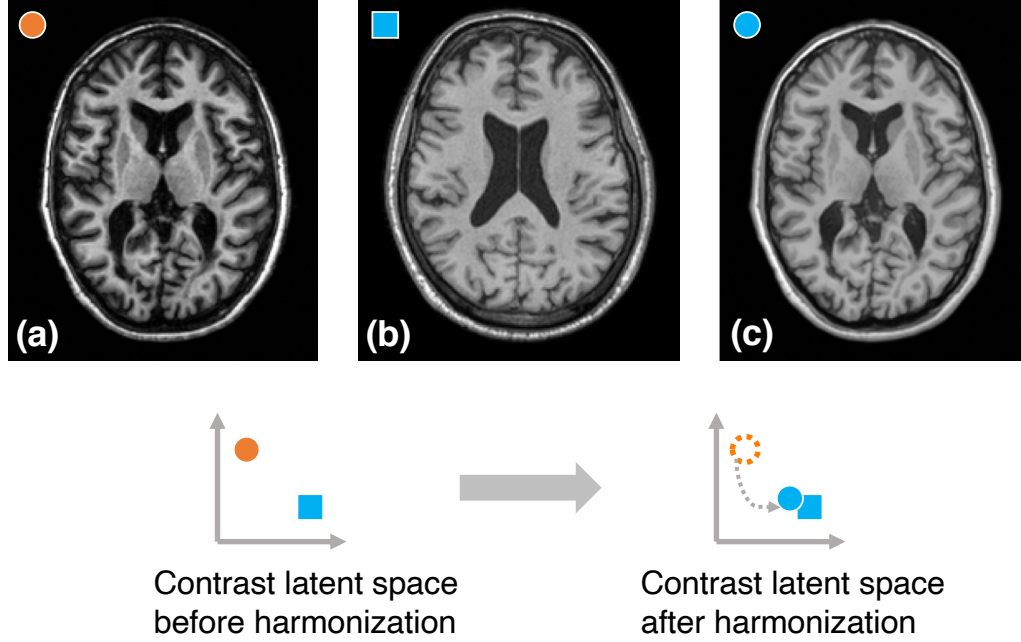
Figure 3-3. Example of CALAMITI [28] result on a T1-w MR image. (a) Source T1-w MR image to be harmonized. (b) Harmonization target T1-w MR image. (c) Harmonized T1-w image after applying CALAMITI on (a). Bottom figure shows illustration of the encoded contrast latent space before and after harmonization. Orange circle represents (a), and blue square and circle represent (b) and (c), respectively.

MR image harmonization algorithm on our segmentation network, demonstrating that image harmonization maintains the anatomy and does not introduce unwanted side effects to the segmentation network. The algorithm provides a potential to process a large amount of data from different sites given a few ground truth labels from one data set.

3.2 Methods

3.2.1 Data Sets and Pre-processing

In this study, we evaluated the performance of the segmentation network using 3D brain T1-w MRI and DTI from 21 subjects. In this data set, all the subjects were diagnosed with

spinocerebellar ataxia (SCA). All the images were acquired on a 3T Philips scanner. The T1-w MPRAGE were acquired with TE = 6 ms, TR = 10.33 ms, FA = 8°, and resolution of $0.82 \times 0.82 \times 1.1$ mm^3. The DWI were acquired using a single shot EPI sequence with one b0 image, 30 gradient directions (b = 700 mm/s^2), minimum TR, TE = 69 ms, FA = 90°, and an acquired spatial resolution of $2.2 \times 2.2 \times 2.2$ mm^3 reconstructed on the scanner to 0.82×0.82 mm^2 axial resolution. Thalamic nuclei masks for the subjects from this data set were created by manual delineation [105] using T1-w MR image, fractional anisotropy image, and Knutsson edge map. The manual delineated mask for each subject includes the following six thalamic nuclei labels: anterior nucleus (ANT), medial dorsal nucleus (MDN), lateral geniculate nucleus (LGN), ventral nuclear group (VNG), pulvinar nucleus (PUL), and medial geniculate nucleus (MGN).

Another data set was used as the harmonization target for the T1-w MR images. The subjects in this data set were diagnosed with mild traumatic brain injury (MTBI). All the images were acquired on a 3T Siemens scanner. The T1-w MPRAGE MRI were acquired with TE = 3.4 ms, TR = 4000 ms, FA = 6°, and resolution of $1 \times 1 \times 1$ mm^3. The DWI were acquired with 6 b0 images, 136 gradient directions (b = 1000 mm/s^2 and 2500 mm/s^2) TE = 78 ms, TR = 3500 ms, FA = 90°, and spatial resolution of $2 \times 2 \times 2$ mm^3. There is no manual delineation of the thalamic nuclei for this MTBI data set.

The T1-w MR images in both data sets were pre-processed by inhomogeneity correction through N4 [67] and rigid registration to MNI 152 atlas space [68]. The pre-processing steps for DWI data include distortion correction [112] and co-registration of the DWI to the anatomical MR images [113]. The diffusion tensor were reconstructed by a non-linear and constrained diffusion tensor fitting algorithm [114] from a software package Camino [115].
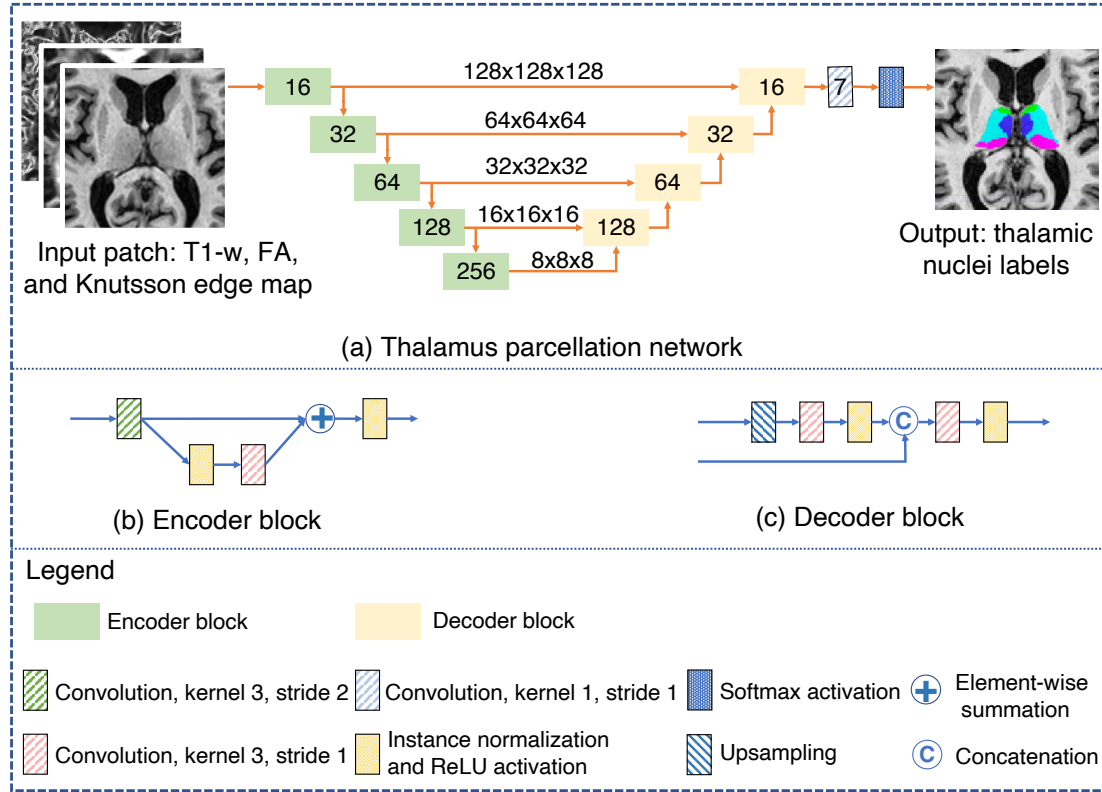
Figure 3-4. (a) Architecture of the thalamus parcellation network. The numbers in the encoder and decoder blocks indicate the number of features. The shape of the feature tensor is denoted at each resolution level. (b) Encoder block. (c) Decoder block.

3.2.2 Thalamus Parcellation Network

A modified 3D U-Net was designed to segment the thalamus and parcellate it into 6 nuclei: anterior nucleus (ANT), medial dorsal nucleus (MDN), ventral nuclear group (VNG), pulvinar nucleus (PUL), medial geniculate nucleus (MGN), and lateral geniculate nucleus (LGN). Figure 3-4 shows the architecture of the thalamus parcellation network. The input to the network is a multi-channel patch including the T1-w MR image, and two feature images derived from DTI, including fractional anisotropy (FA) image and the Knutsson edge maps. Details of the two DTI features are described in Section 3.1.2. All the input images are cropped as $128 \times 128 \times 128$ patches around the image center before input to the network.

Each encoder block consists of a residual unit [71]. The instance normalization layer is used to standardize the features. The final decoder block is followed by a convolutional layer to reduce the number of features to 7, i.e., 6 thalamic nuclei labels and a background label.

During network training, we applied random rotation, scaling, and translation to the training data for data augmentation. The loss function is one minus the mean Dice coefficient of each label (see details in Section 2.2.4).

3.2.3 Image Harmonization and Network Training

In this work, we investigated the impact of the MR image harmonization algorithm, CALAMITI, on our thalamus parcellation network. After the pre-processing steps described in Section 3.2.1, the T1-w MPRAGE images in the SCA data set were further processed in two ways: white matter (WM) peak normalization [116] and CALAMITI (with the MTBI data set as the harmonization target).

We applied 3-fold cross validation on the SCA data set to evaluate the impact of CALAMITI. The 21 subjects were split into 3 groups evenly. During cross validation, we trained the thalamus parcellation network 6 times. All the networks are the same except for their training data. Three of them were trained and tested on unharmonized data, where the T1-w MRIs were only processed by WM peak normalization. The network trained in the ith time, $i \in \{1, 2, 3\}$, took the ith group as testing set and the remaining 2 groups as training set. The other three networks were trained and tested the same way, but on the harmonized data, where the T1-w MRIs were processed by CALAMITI. All the networks were trained for 1000 epochs using the Adam optimizer with initial learning rate of 0.001.
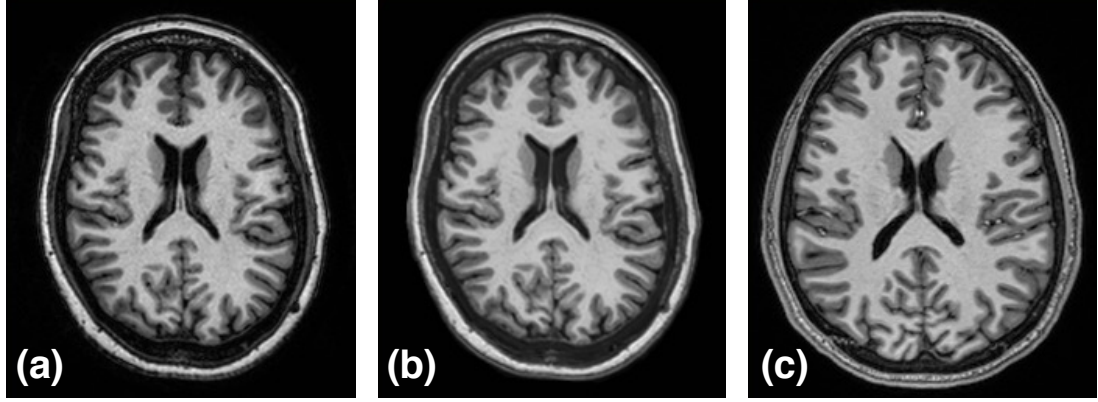
Figure 3-5. Visualization of the T1-w MR image harmonization. (a) An axial slice from a subject in the SCA data set before harmonization. (b) The same axial slice after harmonization. (c) Harmonization target in the MTBI data set.

3.3   Experiments and Results

3.3.1   T1-w Harmonization

A visual comparison of the T1-w MR image before and after CALAMITI is shown in Figure 3-5. It can be observed that the gray matter in the harmonized T1-w image (Figure 3-5(b) is brighter compared to the unharmonized image (Figure 3-5(a)). Figure 3-6 presents the contrast latent space, which is a 2D vector for each T1-w image (see details in Section 3.1.4). In Figure 3-6 left, all the images were processed by WM peak normalization. Since the SCA and the MTBI data sets were acquired under different imaging protocols, the coral circles and the green squares are located in different clusters. After CALAMITI, the latent vectors of the harmonized T1-w MRIs from the SCA data set (gray circles and from the MTBI data set (gray squares) in Figure 3-6 right) group around the harmonization target, which means all the T1-w images from the two data sets now look similar in terms of contrast.
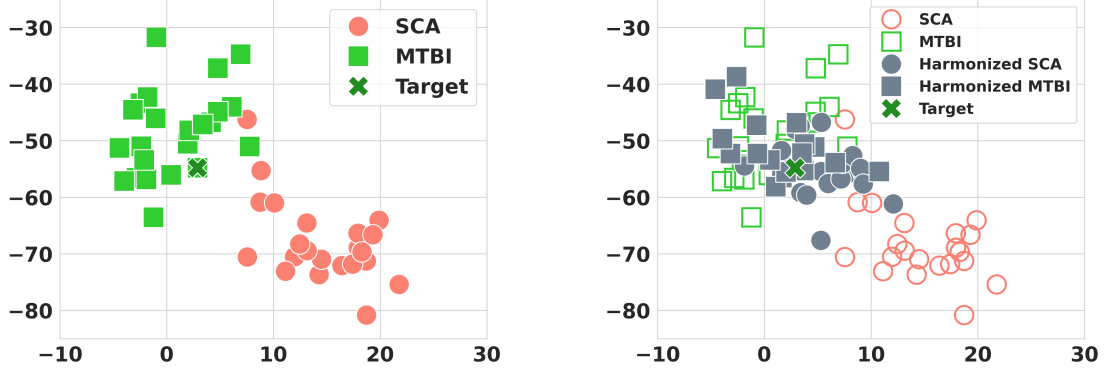
Figure 3-6. Left: A visualization of the contrast latent space (2D vector) of the T1-w images from the SCA and MTBI data sets before harmonization. The "target" cross is the harmonization target in our experiments. Right: A visualization of the contrast latent space of the SCA and MTBI data sets before and after harmonization.

3.3.2 Thalamus Segmentation Comparison

To quantify the performance of the proposed network for whole thalamus segmentation, we compared it with three state-of-the-art brain segmentation methods: FreeSurfer [2, 85], Thalamus Optimized Multi Atlas Segmentation (THOMAS) [95], and Spatially Localized Atlas Network Tiles (SLANT) [117]. In this experiment, we ran each method on the 21 subjects from the SCA data set. We should note that FreeSurfer and THOMAS have different protocols for parcellation of the thalamus. SLANT is a deep learning based whole brain segmentation method. It only produces the whole thalamus mask and does not provide thalamic nuclei segmentation. Therefore, we only compare the performance of the whole thalamus segmentation from each method against manual delineations.

A visual comparison of the thalamus segmentation results produced by each method is shown in Figure 3-7. It can be observed that the networks trained on unharmonized and harmonized data produced similar segmentation results and show better correspondence to the manual delineation compared to other methods. All the methods were quantitatively evaluated by Dice similarity coefficient (DSC) and 95% hausdorff
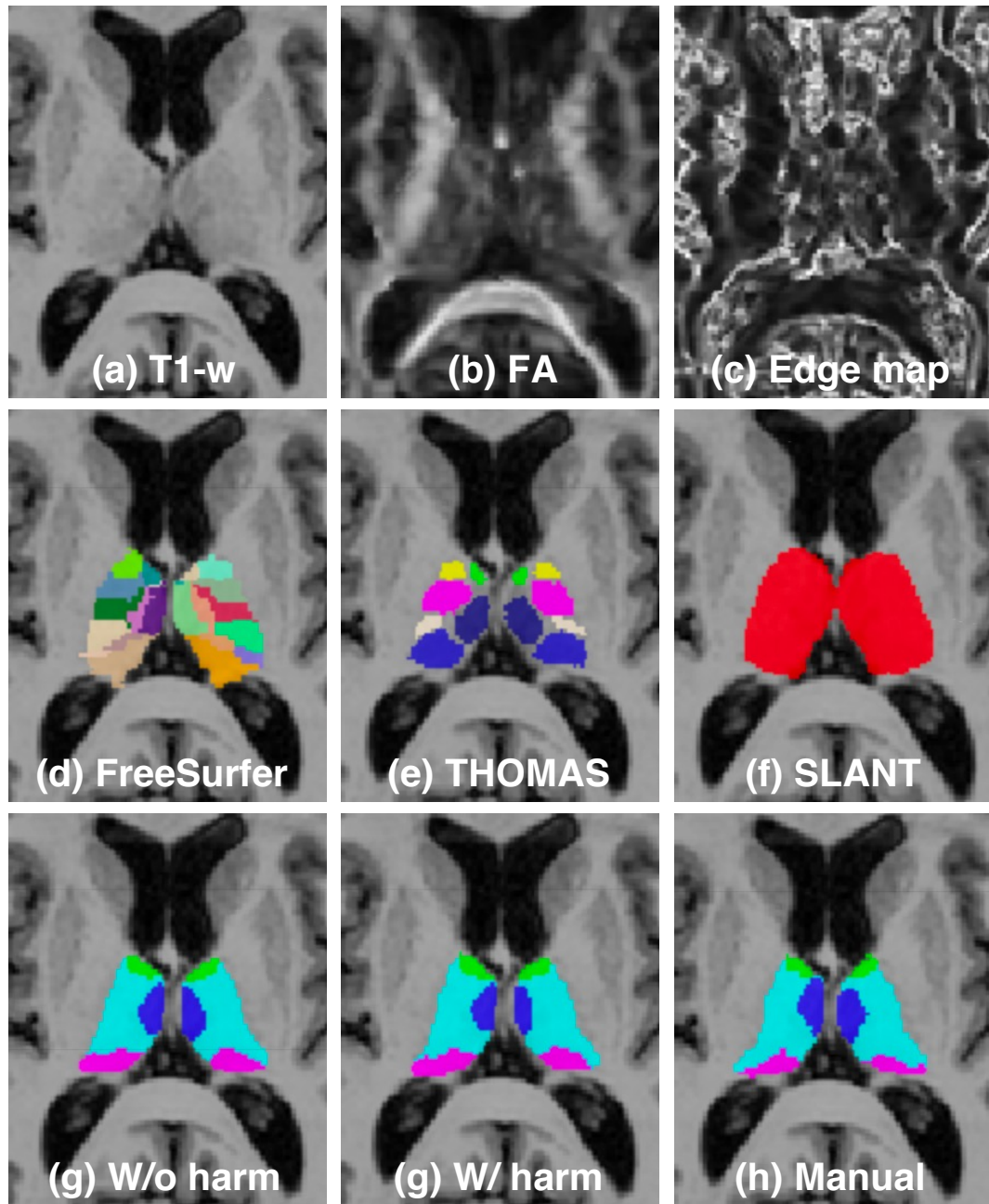
Figure 3-7. Visual comparison of the thalamus segmentation results. (a)–(c): input images to the proposed thalamus parcellation network. (d)–(g): thalamus segmentation or parcellation results produced by different algorithms. (h): manual labels. W/o harm: the proposed thalamic parcellation network trained and tested on the unharmonized data. W/ harm: the proposed thalamic parcellation network trained and tested on the harmonized data.

Table 3-I. The mean±standard deviation of Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD) over 21 subjects from the SCA data set generated by each algorithm.

	DSC	95% HD (mm)
FreeSurfer	0.781±0.029	2.459±0.388
THOMAS	0.680±0.074	3.406±2.416
SLANT	0.757±0.045	2.660±0.466
CNN w/o harmonization	0.853±0.056	1.740±1.107
CNN w/ harmonization	0.861±0.031	1.440±0.386

distance (HD) (see details in Section 2.3.2). The metrics were computed using Segmentaion Metrics Package [118]. Boxplots of the Dice, 95% HD over the 21 subjects from the SCA data set are shown in Figure 3-8. Detailed mean and standard deviation values of each metric are reported in Table 3-I. The proposed networks show higher Dice and lower 95% HD compared ot the other methods on whole thalamus segmentation. The Dice and 95% HD differences are statistically significant in a paired Wilcoxon signed-rank test with an α-level of 0.005. We found no significant differences between the proposed networks trained on unharmonized (CNN w/o harmonization in Table 3-I) or harmonized data (CNN w/ harmonization in Table 3-I). However, it can be observed from Table 3-I that the network trained on harmonized data show lower standard deviations of the two metrics than the network trained on unharmonized data, indicating that harmonization help the network generate more consistent segmentation and reduce outliers.

3.3.3   Thalamic Nuclei Parcellation

Since existing thalamus parcellation algorithms have different thalamic nuclei protocols from our method, we only compare the thalamus parcellation performance between the proposed networks trained on unharmonized or harmonized data against our manual delineations. Boxplots of the DSC and 95% HD are shown in Figure 3-9 and detailed mean and standard deviations of the metrics are reported in Table 3-II. We note that
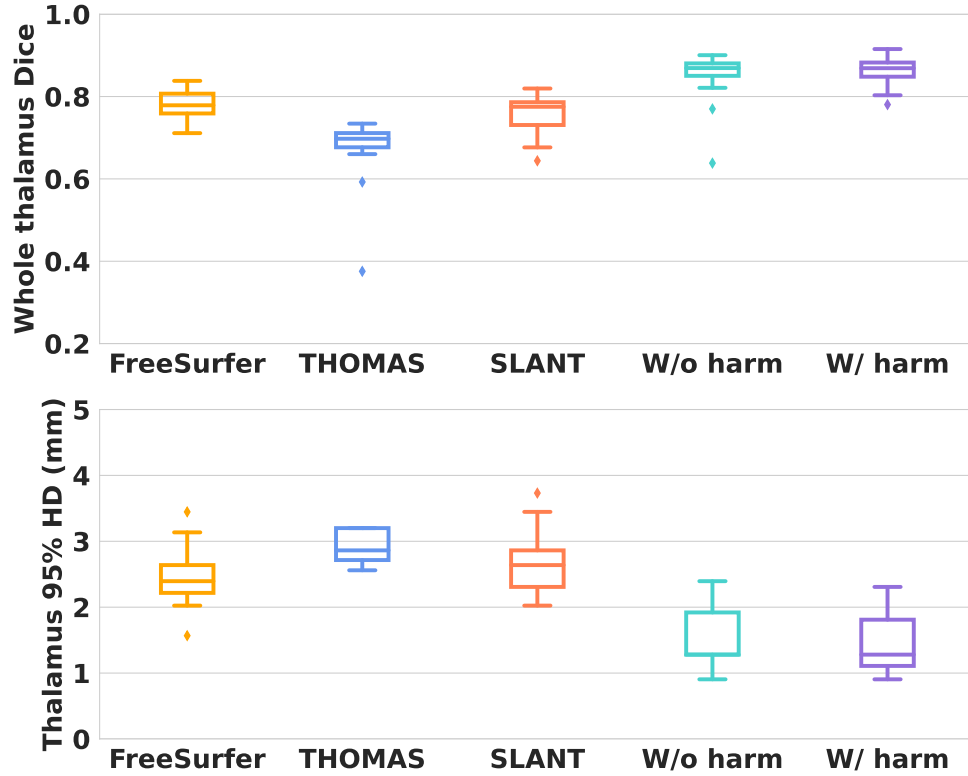
Figure 3-8. Boxplots of Dice similarity coefficients and 95% Hausdorff distance (HD) on the whole thalamus segmentation over 21 T1-w MRIs from the SCA data set. W/o harm: the proposed thalamic parcellation network trained and tested on the unharmonized data. W/ harm: the proposed thalamic parcellation network trained and tested on the harmonized data.

the network trained on harmonized data achieved competitive results compared to the network trained on unharmonized data in terms of DSC and 95% HD. We found no significant differences between the two networks for all thalamic nuclei labels in a paired Wilcoxon signed-rank test. However, for DSC, the network trained on harmonized data achieved higher mean values and lower standard deviations, though not significant, than the network trained on harmonized data except for the ANT label. This is also true for 95% HD. The network trained on harmonized data shows lower 95% HD and lower standard deviations for each nuclei label. The results demonstrate that the harmonization can maintain the underlying anatomy and does not introduce adverse effects to our segmentation network.
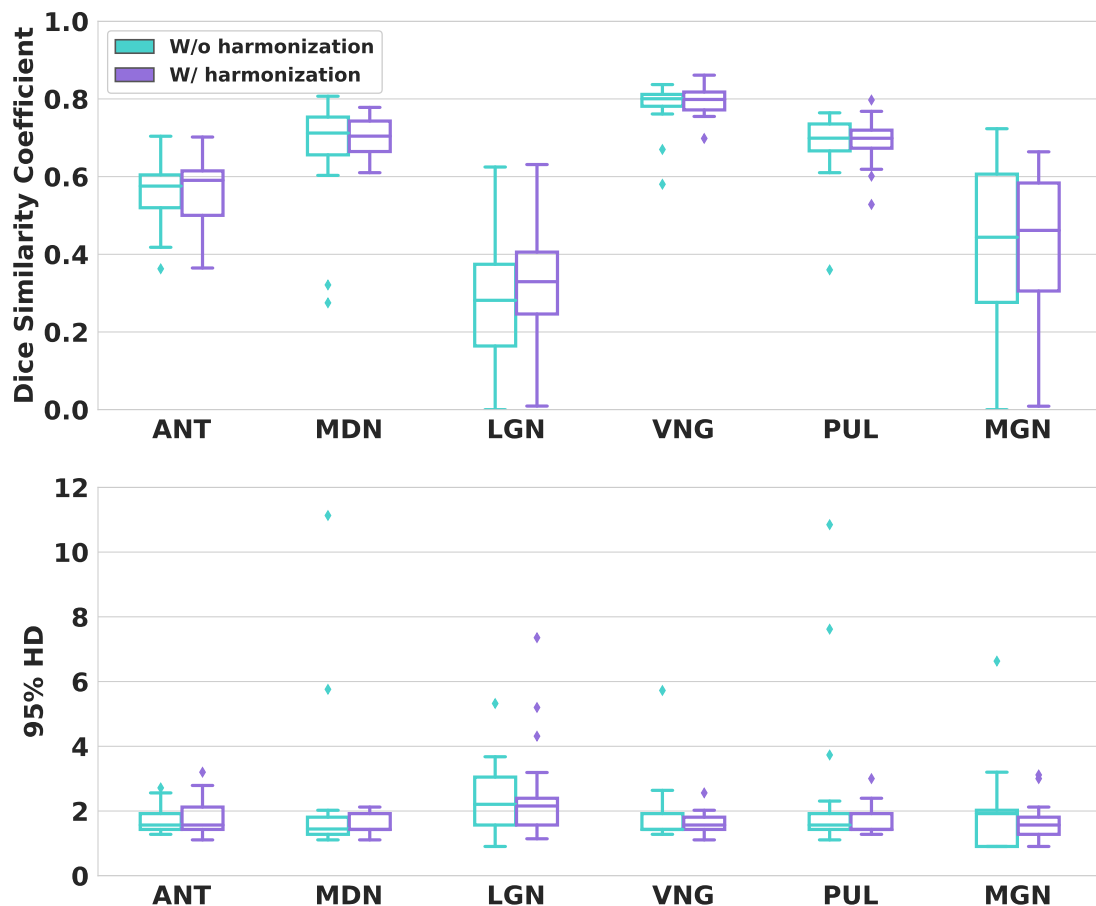
Figure 3-9. Boxplots of Dice similarity coefficients and 95% Hausdorff distance (HD, in mm) on the thalamic nuclei masks produced by the two networks: trained with unharmonized data (W/o harmonization) and trained with harmonized data (W/ harmonization). The testing data are the 21 subjects from the SCA data set. Thalamic nuclei key: anterior nucleus (ANT), medial dorsal nucleus (MDN), lateral geniculate nucleus (LGN), ventral nuclear group (VNG), pulvinar nucleus (PUL), and medial geniculate nucleus (MGN).

3.3.4 Quantitative Evaluation on the MTBI Data Set

The purpose of applying image harmonization to the training data of the segmentation network is to reduce the contrast variation between the training and testing data, therefore enabling the network to process data from different sites. Before harmonization, the T1-w MR images in the MTBI data set have different contrast from those in the SCA data set. After harmonization, all the T1-w MR images were translated to a target image and the contrasts were matched between the two data sets. See Section 3.3.1 for more

Table 3-II. The mean (standard deviation) of Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD, in mm) of thalamic nuclei over 21 subjects from the SCA data set generated by the proposed networks trained on unharmonized data (W/o harm) or harmonized data (W/ harm). Thalamic nuclei key: anterior nucleus (ANT), medial dorsal nucleus (MDN), lateral geniculate nucleus (LGN), ventral nuclear group (VNG), pulvinar nucleus (PUL), and medial geniculate nucleus (MGN).

	ANT	MDN	LGN	VNG	PUL	MGN
DSC:						
W/o harm	0.559(0.086)	0.679(0.135)	0.286(0.168)	0.784(0.056)	0.681(0.083)	0.422(0.202)
W/ harm	0.561(0.102)	0.701(0.050)	0.318(0.153)	0.795(0.034)	0.692(0.059)	0.429(0.176)
95% HD:						
W/o harm	3.989(10.26)	2.143(2.221)	3.714(6.258)	1.819(0.951)	2.391(2.336)	1.916(1.238)
W/ harm	1.828(0.570)	1.545(0.303)	2.544(1.442)	1.614(0.324)	1.687(0.443)	1.644(0.580)

Table 3-III. The mean±standard deviation of Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD) over 94 axial slices from 8 MTBI data generated by the networks.

	DSC	95% HD (mm)
CNN w/o harmonization	0.718±0.167	8.861±8.537
CNN w/ harmonization	0.875±0.095	2.328±1.041

details about the harmonization processing.

We evaluated the performance of the networks described in Section 3.2.3 on the unharmonized and harmonized MTBI data, respectively. We manually delineated the whole thalamus of some key axial slices in 8 MTBI subjects, and there are 94 slices with manual mask in total. Figure 3-10 presents the boxplots of the DSC and 95% HD on the whole thalamus masks for the MTBI data produced by the networks. Detailed values of mean and standard deviation of the two metrics are reported in Table 3-III. The network trained with harmonized data showed significant improvement ($p < 0.005$) over the network trained with unharmonized data based on a paired Wilcoxon signed-rank test. Therefore, the network trained with harmonized SCA data can be applied on data from the MTBI data set without creating new manual labels or retraining the network on the new data set.
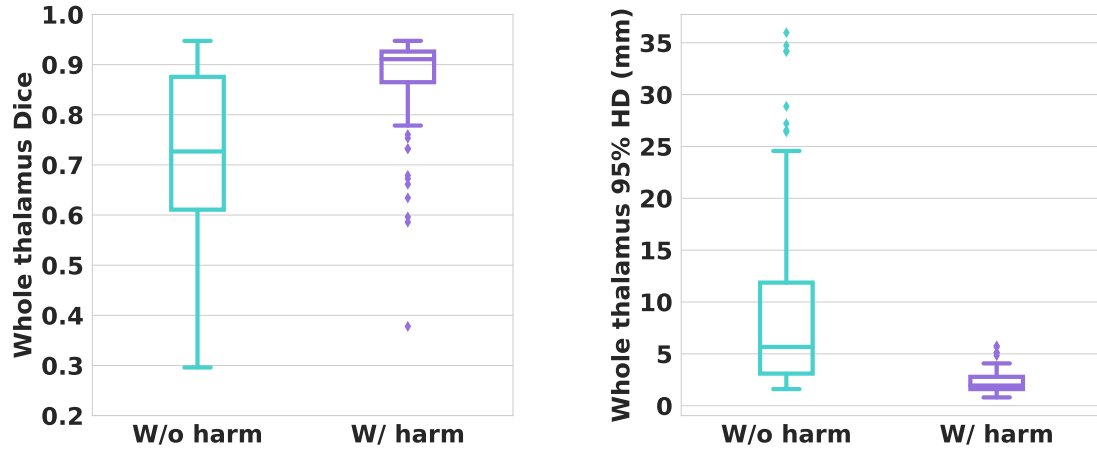
Figure 3-10. Boxplots of Dice similarity coefficients and 95% Hausdorff distance on the whole thalamus segmentation over 94 2D slices from the MTBI data set. W/o harm: the proposed thalamic parcellation network trained and tested on the unharmonized data. W/ harm: the proposed thalamic parcellation network trained and tested on the harmonized data. The Dice and 95% HD were calculated on select 2D axial slices, as we do not have a complete 3D delineation. See Sec. 3.3.4 for more details.

3.4 Discussion

The proposed thalamus parcellation network incorporates features derived from DTI, which provides better contrast of the thalamus boundary than structural MRIs. We have performed evaluation of the proposed thalamus parcellation network on two different data sets, i.e. SCA and MTBI. The accuracy of the segmentation was evaluated using DSC and 95% HD between the automated parcellation and manual delineations. In the experiment, the proposed network shows higher DSC and lower 95% HD on the whole thalamus segmentation compared to widely used thalamus segmentation algorithms, as shown in Table 3-I and Figure 3-8.

We investigated the impact of an MR image harmonization algorithm, CALAMITI, on the proposed network. The network trained on T1-w images harmonized to a new target achieved competitive performance comparing to the network trained on unharmonized images, demonstrating that the harmonization can maintain the underlying anatomy and

does not introduce adverse effects to the proposed network. Comparing the performance of the networks trained on different data, the network trained on harmonized data achieved lower standard deviations than the network trained on unharmonized data, indicating that harmonization help the network generate more consistent segmentation and reduce outliers. When the networks trained on different data were applied to the harmonization target data set, which are unseen data for both networks, the network trained on harmonized data showed significantly better segmentation results compared to the other one. Therefore, the harmonization algorithm provides potential to process large amounts of data from different sites. The new testing data can be first harmonized to the same target data and then input to the trained network, which will reduce the contrast variation between the training and testing data and hence ease the domain shift problem.

3.5 Summary

In this chapter, we presented a multi-channel 3D U-Net for thalamus parcellation using multimodal MRIs. The proposed network utilizes T1-w MR image and features derived from DTI. We showed that proposed network achieves better performance on the whole thalamus segmentation than the comparing methods. We investigated the impact of an MR image harmonization algorithm on the proposed network. The network trained on harmonized data showed competitive results when evaluated on the same data set and significantly better results on an unseen data set. The experiments demonstrated that the harmonization can adjust the image contrast while maintaining the under anatomy. Therefore, the network trained on harmonized data provides the potential to process large amounts of data from other data sets, even in the absence of site-specific training data.

Chapter 4

Reconstruction of Tongue Muscle Fiber Orientations using Diffusion MRI

4.1 Introduction

4.1.1 Anatomy of the Human Tongue

The tongue plays an important role in multiple vital human functions including breathing, swallowing, and speaking [119]. These functions are possible because the tongue can produce complex deformations with its muscle architecture. This muscle architecture consists of intrinsic and extrinsic muscle groups. The transversalis (T), verticalis (V), superior (SL) and inferior longitudinalis (IL) muscles are four intrinsic muscle groups contained within the tongue. They are responsible for altering the shape of the tongue, such as lengthening, shortening, curling and flattening. The genioglossus (GG), hyoglossus (HG), styloglossus (SG), and geniohyoid (GH) are four extrinsic muscle groups that arise from the external organs and are inserted into the tongue body [120]. These muscles are capable of moving the tongue in different directions to alter the tongue's position. The intrinsic and extrinsic muscles cooperate in an elaborate fashion to create

64

various deformations during tongue motion. Figure 4-1 shows the anatomy of the tongue with muscular structures in a sagittal section of the head [121].
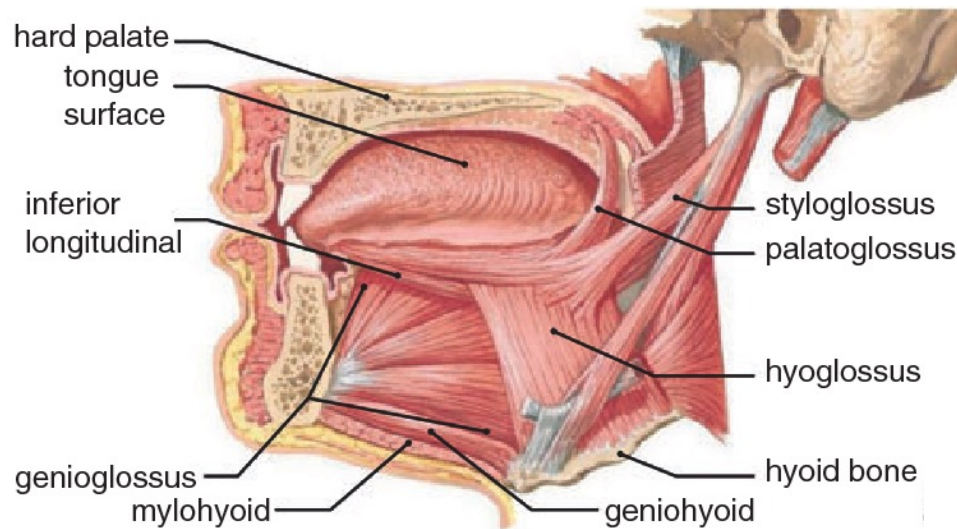
Figure 4-1. Sagittal section of the head, showing the tongue and its muscular structures [121].

One important characteristic of the tongue muscles is that the tongue muscle fibers are extensively interdigitated and nearly orthogonal in three dimensions [119, 122]. In the tongue, some regions contain a single fiber orientation, but most regions are interdigitated with two fiber orientations crossing at approximately 90 degrees. Figure 4-2 shows a T2-weighted MRI of a human tongue and idealized tongue muscle fiber orientations from a fiber orientation model [123]. The fiber orientations are displayed in two figures, which are denoted as first and second fiber orientation images. There are compelling scientific reasons to obtain a better understanding of the relationship between individual tongue muscles and the complex functions [30]. For example, it is essential to obtain the muscular architecture to better model and predict the functional outcome after surgical removal of part of the tongue.
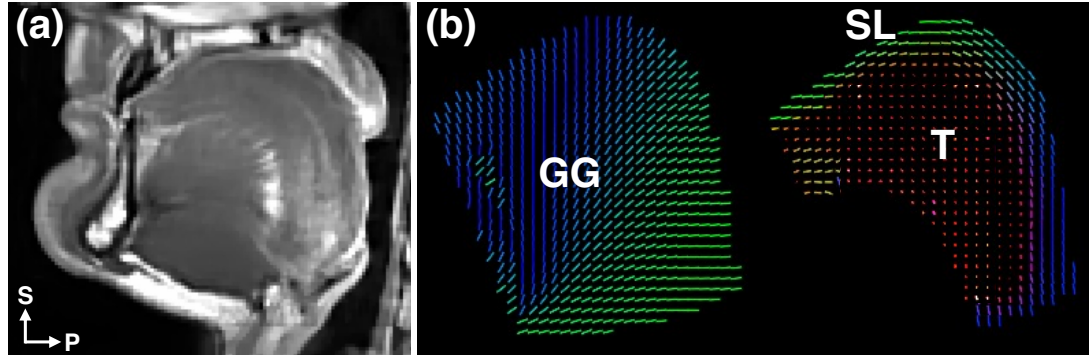
Figure 4-2. (a) Mid-sagittal slice of a T2-w MRI of a human tongue. S: superior; P: posterior. (b) Sagittal view of the tongue muscle fiber orientations generated based on a mechanical tongue model [123]. The muscle groups in (b) are: genioglossus (GG), superior longitudinalis (SL), and transversalis (T). The fiber orientations are conventionally color-coded (red: right-left, green: anterior-posterior, blue: inferior-superior).

4.1.2 High Angular Resolution Diffusion Imaging

Diffusion weighted imaging (DWI) is an MRI modality that measures the diffusion of water in tissues, as described in Section 3.1.2. It has been used to image tongue muscle architecture and interpret the muscles' functions [30, 124, 125]. The commonly used DWI technique, diffusion tensor imaging (DTI), can only describe a single fiber orientation within each voxel, i.e., the major eigenvector of the diffusion tensor [126]. This is not adequate for tongue fiber orientation reconstruction because of the highly interdigitated muscles in the tongue. To discriminate crossing fiber orientations in the tongue, high angular resolution diffusion imaging (HARDI) [127] can be used.

The HARDI technique consists of dense sampling of the signal on a sphere in the 3D diffusion space, also called q-space. It acquires diffusion data with a large number of gradient directions, typically more than 45 [102], at one or more fixed levels of diffusion weighting, i.e., the b-values. If only a single b-value is used in the acquisition, the sampling scheme is generally called single-shell HARDI. If multiple b-values are used, it is called multi-shell HARDI. In this chapter, we will focus on single-shell HARDI data.
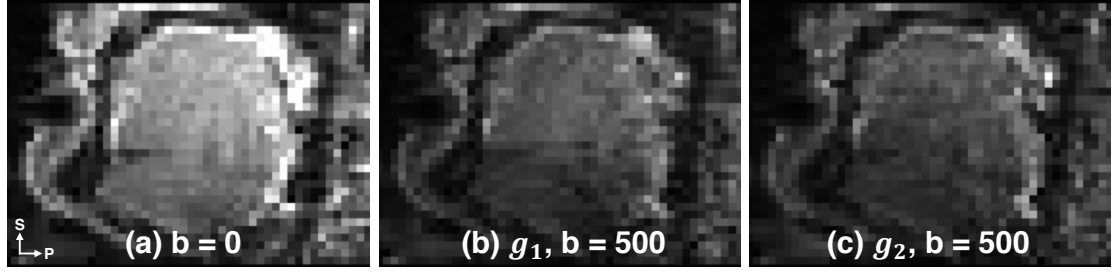
Figure 4-3. Examples of HARDI data from sagittal slices through the same subject's tongue, including (a) a b0 image, and (b, c) DWIs with b-value of 500 and two different gradient directions g. S: superior; P: posterior.

With the high angular resolution, HARDI provides a more accurate representation of the motion pattern of water diffusion within a voxel. Figure 4-3 shows examples of HARDI tongue data with different gradient directions g.

At each imaging voxel of the HARDI data, the signal is sampled along different gradient directions on the surface of a sphere. Therefore, the HARDI signal s can be represented using spherical harmonics (SHs) [128]. Spherical harmonics of order l and phase m, denoted as $Y_l^m(\theta, \phi)$ or Y_{lm} [129], form an orthonormal basis defined on a sphere, where θ and ϕ are the spherical angular coordinates. Figure 4-4 shows visual representations of the first few spherical harmonics. Any spherical function can be decomposed as a linear combination of these SHs, which is similar to the Fourier transform in the plane, but defined on the domain of the sphere.

The HARDI signal S is real and symmetric due to the nature of the diffusion MRI signal, where the diffusion encoding is incapable of distinguishing between diffusion along or opposite the gradient direction. Therefore, it is common to only use even order terms of the SHs. To compute the SH representation with order $L = l_{\max}$ of a HARDI signal, the total number of SH coefficients need to be estimated is $R = (L+1)(L+2)/2$. The raw HARDI signal S at gradient direction g_i is estimated as:
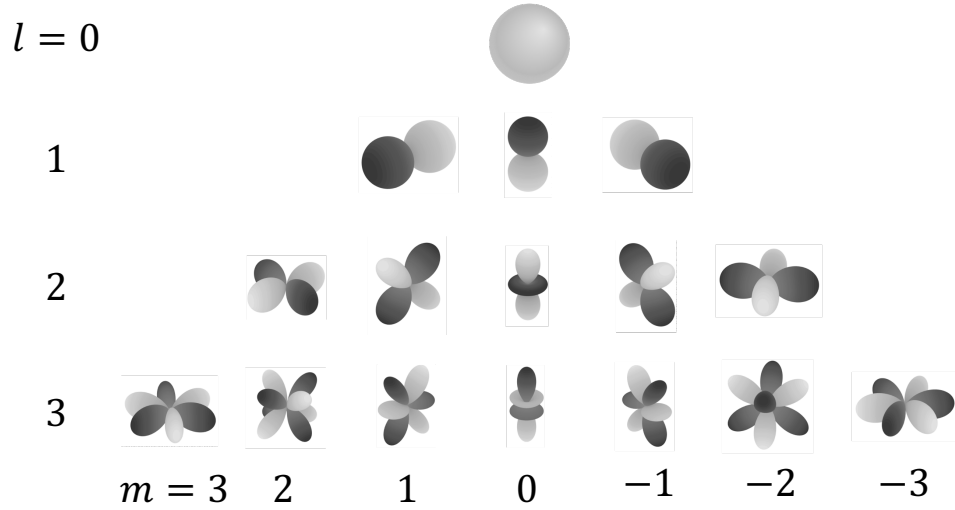
Figure 4-4. Visual representations [130] of the first few real spherical harmonics Y_{lm} for $l = 0, ..., 3$ (top to bottom) and $m = -l, ..., l$ (left to right).

$$S(\boldsymbol{g}_i) = \sum_{j=1}^{R} c_j Y_j(\boldsymbol{g}_i), i = 1, ..., N \tag{4.1}$$

where $\{c_j\}$ are set of SH coefficients need to fit and N is the total number of gradient directions during image acquisition. The HARDI signal at each voxel along all the gradient directions can be written in a matrix form:

$$\boldsymbol{s} = \boldsymbol{cB} \tag{4.2}$$

where $\boldsymbol{s} = [S(\boldsymbol{g}_1), ..., S(\boldsymbol{g}_N)]$, $\boldsymbol{c} = [c_1, ..., c_R]$, and $\boldsymbol{B}$ is the $R \times N$ basis matrix of the SHs evaluated at each gradient direction $\boldsymbol{g}_i$. The SH coefficients $\boldsymbol{c}$ can be estimated by least-squares approach [131]. One property of spherical harmonics is that it maps the raw signal to the spherical space and represents the data with the same number of coefficients, which is convenient for further analysis.

4.1.3 Related Work

Previous studies have investigated techniques to reconstruct the crossing fiber orientations based on HARDI data. Tuch et al. [132] proposed a framework to resolve intravoxel crossing fibers using diffusion orientation distribution function. Tournier et al. [133, 134] proposed the constrained spherical deconvolution (CSD) algorithm. Spherical convolution assumes that the measured diffusion signal is the convolution over the unit sphere of the underlying fiber orientation distribution function with a response function measured from a single coherently oriented fiber population [133]. Therefore, the fiber orientation distribution function can be solved by spherical deconvolution technique. Aganj et al. [135] proposed a framework to correct a mathematical inaccuracy in the original diffusion orientation distribution model in [132] by considering the solid angle factor to implicitly sharpen the fiber peaks. Behrens et al. [136] proposed a method called Bayesian estimation of diffusion parameters obtained using sampling techniques for crossing fibers (BEDPOSTX). The method used a Markov Chain Monte Carlo sampling scheme to build up distributions on diffusion parameters at each voxel, which can be used for probabilistic tractography. A recently proposed deep learning method [137] used a convolutional neural network to estimate the fiber orientation distribution function. However, all existing fiber orientation reconstruction algorithms prior to our work were designed for brain tissue and do not account for the orthogonal nature of the crossing muscle fibers in the human tongue.

4.1.4 Contribution

In this chapter, we present a deep convolutional neural network to directly reconstruct the crossing muscle fiber orientations from HARDI tongue data in a fully automatic manner [29]. Neighboring information is incorporated using convolutional layers to make the fibers spatially consistent. We introduce a separation loss during training

to encourage orthogonality of the predicted crossing fibers, which is critical in tongue muscle fiber orientation reconstruction. We also present a fiber matching algorithm [31] to refine the assignments of the reconstructed fiber orientation and yield more consistent tongue muscle fiber arrangements.

4.2 Methods

4.2.1 Data Sets and Pre-processing

Since it is nearly impossible to manually annotate the ground truth fiber orientations from the diffusion data, we used synthetic data to train our fiber orientation reconstruction network. The training data were created based on [138]. In this data set, the ground truth fiber orientations and the corresponding volume fractions were generated by applying the multi-fiber ball-and-stick method [139] on the brain of a healthy subject. In our experiments, up to two fibers per voxel were chosen to form the ground truth and synthesize the diffusion data, which is consistent with the tongue anatomy. The fiber orientations and fractions formed the fiber orientation distribution function at each voxel. The synthetic DW images were created using the multi-tensor model [127]. The diffusion-weighting b-value was set to 1000 s/mm^2. We created different gradient schemes with 60, 90, or 120 gradient directions. In each gradient scheme, the gradient directions are evenly distributed on the unit sphere. In each image, Rician noise was added to the synthesized diffusion signal. The signal-to-noise ratio (SNR) of the synthetic DWI was set to 9, 18, or 36 in different images.

The length of the HARDI signal vector s is determined by the number of gradient directions, which is different among imaging protocols. To ensure that the network can be applied to HARDI data with different numbers of gradient directions, the raw HARDI signal was transformed to the spherical harmonic (SH) space. As described
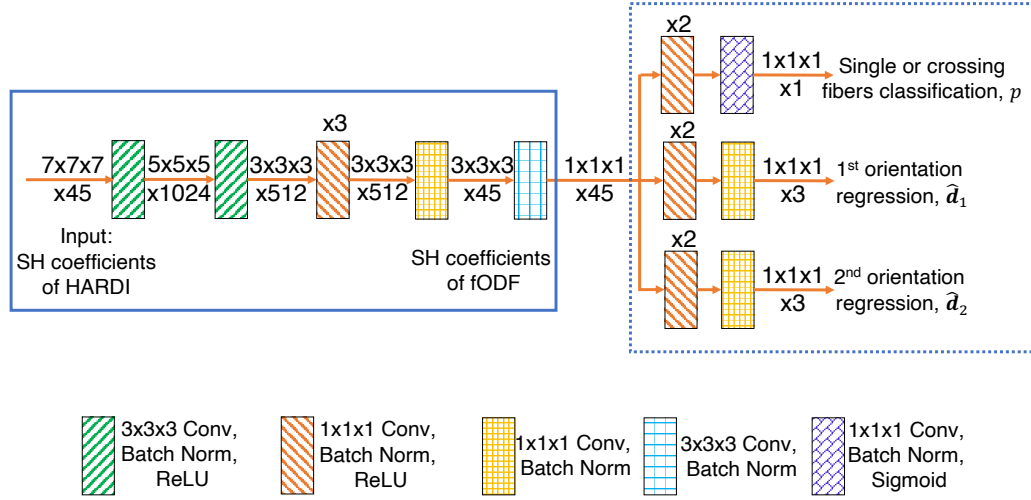
Figure 4-5. Architecture of the proposed fiber estimation network. The first stage (solid box) of the network predicts the SH coefficients of the fODF. The second stage (dots and dashes box) predicts the number of the fibers and the corresponding orientations. The numbers above the orange arrows indicate the shape of the tensor and the numbers below the arrows indicate the numbers of features. The numbers above the colored block indicate repetitions of the block.

in Section 4.1.2, the HARDI signal can be decomposed as a linear combination of the SH basis with even orders. The SH coefficients were computed by least-squares linear fitting to the HARDI signal, as in [134].

4.2.2   Fiber Orientation Reconstruction Network

Given HARDI data of a human tongue, our goal is to estimate both the number of muscle fiber orientations (single or crossing) and the corresponding fiber orientations at each voxel. The input to the network is defined as the SH coefficients of the HARDI signal, which is described in Section 4.2.1. The output at each voxel is divided into two parts. The first part is the SH coefficients of the fiber orientation distribution function (fODF). The second part contains one binary term indicating if the voxel contains a single or crossing fibers and two unit vectors representing the fiber orientations.

Figure 4-5 shows the architecture of the proposed fiber orientation reconstruction

network. It is composed of two stages. The first stage (solid blue box) predicts the SH coefficients of the fODF, and the second one (dots and dashes blue box) predicts the fiber orientations, which simulates the peak finding process in HARDI reconstruction methods. In our experiments, both input HARDI signal and fODF were fitted to 8th order of SHs, which contains 45 SH coefficients.

In the first stage, the input $7 \times 7 \times 7 \times 45$ patch is passed through two $3 \times 3 \times 3$ convolutional blocks and four $1 \times 1 \times 1$ convolutional blocks with the number of features per block being: 1024, 512, 512, 256, 512, and 45 (output). The $1 \times 1 \times 1$ convolutional blocks are used to add more features to the network without decreasing the size of the patches. The output is a $3 \times 3 \times 3 \times 45$ patch representing the SH coefficients of the fODF. This patch is passed through a $3 \times 3 \times 3$ convolutional layer, a batch normalization layer and becomes the input to the second stage, which is the SH coefficients of the fODF in the central voxel.

The second stage consists of three parallel branches, including one classifier that predicts the number of fibers (single or crossing) in the central voxel and two regressors that predict the corresponding fiber orientations in the central voxel. Each part consists of three convolutional blocks. Non-zero output vectors from the regressors are normalized as unit vectors to represent the fiber orientations. In the network, each convolutional block has a convolutional layer, followed by a batch normalization [9] layer and a rectified linear unit (ReLU) [7] layer. ReLU is not used in the last two convolutional blocks of the first stage and the last block of the two regressors in the second stage, since the SH coefficients and the fiber orientations can be negative. The final layer of the classifier in the second stage is activated by a sigmoid function to obtain a probability that predicts if the central voxel has single fiber or crossing fibers. If the probability is greater than 0.5, it indicates the voxel has crossing fibers.

4.2.3 Loss Function and Network Training

The loss function of the fiber orientation reconstruction network is a combination of the mean squared error in the first stage (s1) and the fiber errors in the second stage (s2). The loss in the first stage is given by:

$$\text{Loss}_{\text{s1}} = \frac{1}{R}\sum_{j=1}^{R}(c_j - \hat{c}_j)^2 \tag{4.3}$$

where c_j and $\hat{c}_j$ are the jth true and estimated SH coefficients of the fODF, respectively. R is the total number of the SH coefficients, which is 45 in our experiments. The loss in the second stage is given by:

$$\text{Loss}_{\text{s2}} = l_{\text{CE}}(p, \mathbb{1}) + l_{\text{FOE}}(\hat{\boldsymbol{d}}_1, \boldsymbol{d}_1) + l_{\text{FOE}}(\hat{\boldsymbol{d}}_2, \boldsymbol{d}_2) + l_{\text{sep}}(\hat{\boldsymbol{d}}_1, \hat{\boldsymbol{d}}_2) \tag{4.4}$$

where $l_{\text{CE}}(p, \mathbb{1})$ is the cross entropy loss between the predicted probability of crossing fibers p and the ground truth and is defined as: $l_{\text{CE}}(p, \mathbb{1}) = x\log p + (1 - x)\log(1 - p)$, where x is the ground truth label indicating if the voxel has crossing fibers. The fiber orientation error, denoted as $l_{\text{FOE}}(\hat{\boldsymbol{d}}_m, \boldsymbol{d}_m), m \in \{1, 2\}$, is calculated between the mth predicted fiber orientation $\hat{\boldsymbol{d}}_m$ and the truth orientation $\boldsymbol{d}_m$. l_{FOE} is modified from the cosine similarity loss and is defined as:

$$l_{\text{FOE}}(\hat{\boldsymbol{d}}_m, \boldsymbol{d}_m) = -\frac{|\hat{\boldsymbol{d}}_m \cdot \boldsymbol{d}_m| + \epsilon}{|\hat{\boldsymbol{d}}_m| + |\boldsymbol{d}_m| + \epsilon}, m \in \{1, 2\} \tag{4.5}$$

where ϵ is a small positive number to avoid a zero denominator. The predicted fiber orientation $\hat{\boldsymbol{d}}_m$ is forced to be close to $\boldsymbol{d}_m$ to minimize l_{FOE}. The absolute value of the inner product is taken because two opposite directions are considered to be the same orientation, as the diffusion encoding cannot distinguish between diffusion along or opposite the gradient direction.

To encourage the orthogonality of the crossing fibers in the tongue [122], we propose

a separation loss l_{sep} between the predicted crossing fiber orientations $\hat{\boldsymbol{d}}_1$ and $\hat{\boldsymbol{d}}_2$, which is defined as:

$$l_{\text{sep}}(\hat{\boldsymbol{d}}_1, \hat{\boldsymbol{d}}_2) = |\hat{\boldsymbol{d}}_1 \cdot \hat{\boldsymbol{d}}_2| \tag{4.6}$$

This separation loss at each voxel will be minimized when the crossing angle of the two predicted orientations is 90°. Therefore, the network is encouraged to predict orthogonal crossing fiber orientations, which is an important characteristic in tongue muscles.

The proposed fiber orientation reconstruction network was trained on 5 synthetic HARDI data sets (see Section 4.2.1 for details). Each image contains around 80k patches and the batch size was 1024. Another synthetic data set was used for validation. The network was trained using the Adam optimizer [14] with initial learning rate of 0.0001. The performance of the trained network was evaluated at 150 epochs as the validation loss plateaued.

4.2.4 Fiber Orientation Matching Algorithm

In Section 4.2.2, we presented a deep network to reconstruct the fiber orientations from the HARDI tongue data. The fiber orientation reconstruction network predicts up to two unit vectors per voxel that represent the fiber orientations. When there are two crossing fibers, the one that is assigned to be the primary one, which we referred to as $\boldsymbol{d}_1$, is intended to be more dominant in that voxel. When there is only one fiber, it is always assigned to $\boldsymbol{d}_1$. However, the assignment in one voxel might disagree with a neighboring voxel (see Figure 4-6 bottom). This can happen because of either noisy data or when the diffusivities of the two orientations are similar. Since we want to study the patterns amongst tongue muscle groups, it is important that the assignments of the muscle fiber orientations are smoothly and consistently defined throughout the tongue. In the following sections, we present a fiber matching algorithm to produce more consistent fiber orientations given the direct output from fiber orientation reconstruction algorithms.

This algorithm can be applied to any fiber reconstruction method that estimates one or two fiber orientations per voxel such as the constrained spherical deconvolution [133, 134] algorithm.

4.2.5 Fiber Matching Theory

The input to the fiber matching algorithm is a pair of unit vector images that represent the fiber orientations generated from the fiber reconstruction network. In the voxels with crossing fiber directions, the matching algorithm decides whether to leave the assignment of d_1 and d_2 as it is or switch the assignments between d_1 and d_2. In the voxels with a single fiber direction, the matching algorithm decides whether to leave d_1 as it is or to reassign it to d_2.

In each voxel, there are two fiber orientations d_1 and d_2. In the single fiber region, d_2 is defined as a zero vector. A binary variable y is introduced in each voxel to indicate if the current assignment is correct. In our experiments, $y = 1$ means the current assignment is correct, and $y = 0$ means the d_1 and d_2 assignment should be switched to be more consistent with its neighboring voxels. We consider two neighboring voxels p and q and the two orientations at each voxel: $d_1(p)$, $d_2(p)$, $d_1(q)$, and $d_2(q)$. The following quantities measure the similarity between orientations in voxels p and q under different assignments:

$$a_{11} = |d_1(p) \cdot d_1(q)|$$
$$a_{12} = |d_1(p) \cdot d_2(q)|$$
$$a_{21} = |d_2(p) \cdot d_1(q)|$$
$$a_{22} = |d_2(p) \cdot d_2(q)| \tag{4.7}$$

where a_{mn} is the absolute value of the cosine similarity between the mth orientation in

voxel p and the nth orientation in voxel q. Greater a_{mn} indicates larger similarity. Figure 4-6 illustrates the correct and incorrect assignments of the crossing fiber orientations in two neighboring voxels p and q in ideal case, where the two voxels have the same crossing fiber orientations. The blue and orange lines represent the first and second fiber orientations, respectively. In case where the assignments are correct (Figure 4-6 top), the similarity measurement a_{11} is the absolute value of the inner product of the two blue lines, which is 1. The similarity measurement a_{12} is the absolute value of the inner product of the blue line in voxel p and orange line in voxel q, which is 0. The same calculation can be done for a_{21} and a_{22}. It can be observed that in the correct assignment case, a_{11} and a_{22} are large and a_{12} and a_{21} are small. Turning to the incorrect assignment case (Figure 4-6 bottom), a_{11} is still the absolute value of the inner product of the two blue lines, but the value is 0. The similarity measurement a_{12} is 1 in this case. Hence, in the incorrect assignment case, a_{12} and a_{21} are large and a_{11} and a_{22} are small.

4.2.6   Implementation of the Fiber Orientation Matching Algorithm

We define the fiber orientations in voxel p and q after correction as: $\boldsymbol{d}_1^*(p)$, $\boldsymbol{d}_2^*(p)$, $\boldsymbol{d}_1^*(q)$, $\boldsymbol{d}_2^*(q)$. The refinement of the orientations can be formed as an optimization problem. The objective function for all voxels in the tongue is formed as

$$\max \sum_{(y_p, y_q) \in \mathcal{E}} \left(|\boldsymbol{d}_1^*(p) \cdot \boldsymbol{d}_1^*(q)| + |\boldsymbol{d}_2^*(p) \cdot \boldsymbol{d}_2^*(q)| \right) \tag{4.8}$$

where $\mathcal{E}$ is the set of all neighboring voxel pairs in the image. The matching algorithm is designed to find the best assignment in each voxel that will maximize the similarity between the first orientation in that voxel and its neighboring voxels and the similarity between the second orientation in that voxel and its neighboring voxels. After the optimization, the display of the fiber orientations will be more consistent. In each voxel,

76
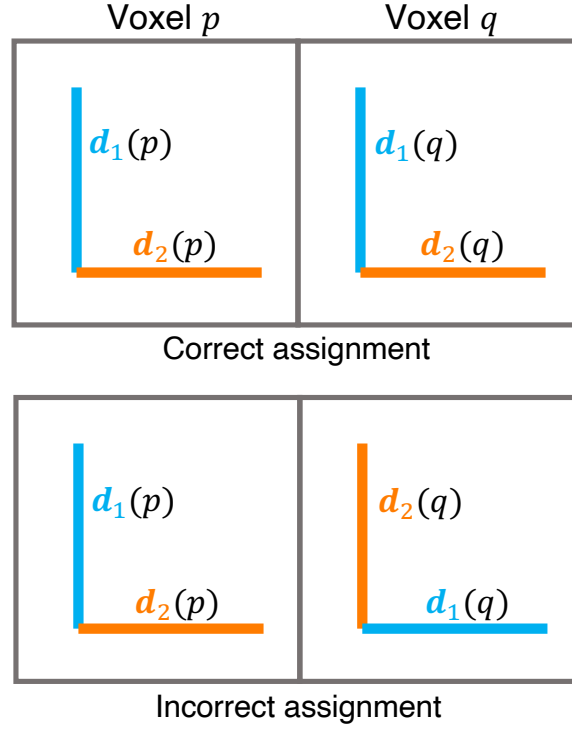
Figure 4-6. Illustration of the correct (top) and incorrect (bottom) assignments of two neighboring voxles p and q. Blue and orange lines represent the first and second fiber orientations, respectively.

there are 4 combinations of assignment pair: $(y_p, y_q) \in \{(1,1), (1,0), (0,1), (0,0)\}$.

Equation 4.8 can be expanded as a function of y and a_{mn}:

$$\max \sum_{(y_p, y_q) \in \mathcal{E}} y_p y_q (a_{11} + a_{22}) + y_p (1 - y_q)(a_{12} + a_{21}) +$$

$$(1 - y_p) y_q (a_{21} + a_{12}) + (1 - y_p)(1 - y_q)(a_{22} + a_{11}) \qquad (4.9)$$

and simplified as

$$\max \sum_{(y_p, y_q) \in \mathcal{E}} [y_p y_q (2a_{11} + 2a_{22} - 2a_{12} - 2a_{21}) +$$

$$y_p(a_{12} + a_{21} - a_{11} - a_{22}) +$$

$$y_q(a_{21} + a_{12} - a_{11} - a_{22})] \tag{4.10}$$

which ignores the constant term $(a_{11} + a_{22})$. This optimization can be modeled as a quadratic unconstrained binary optimization (QUBO) [140] problem, which is an unconstrained model, other than the binary restrictions on the decision variables y_p and y_q. The QUBO problem is NP-hard and the solution can be found iteratively by partitioning it into pieces that can be solved by a quantum annealing system [141]. Solving this QUBO model gives y in each voxel to determine if the assignment of $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ should be switched.

4.3   Experiment and Results

4.3.1   Experimental Setup for Fiber Orientation Reconstruction

Comparison Methods

The performance of the proposed fiber orientation reconstruction network was compared with two widely used fiber reconstruction methods: constrained spherical deconvolution (CSD) [134], and Bayesian estimation of diffusion parameters obtained using sampling techniques (BEDPOSTX) [136, 139]. CSD was implemented in Dipy [142], and BEDPOSTX was implemented in FMRIB Software Library (FSL, version 6.0) [143]. When running CSD, a response function calculated from a single fiber bundle is required. We manually selected the region of interest in the genioglossus (GG) muscle in a known

non-crossing area to estimate the response function. The output of CSD is the fODF, where a peak finding algorithm [134] was performed to identify distinct fiber orientations. A relative peak threshold was used to filter out some small peaks introduced by noise. In this study, we set this threshold to 0.2, as recommended in [144]. When running BEDPOSTX, we thresholded the anisotropic volume fraction at 0.01 to remove spurious peaks. The default threshold in FSL is 0.05. However, BEDPOSTX identified the number of voxels contain crossing fibers as 4% in a synthetic HARDI tongue data using the default threshold, when the true number of crossing fiber voxels is 67%. Therefore, we lower the volume fraction threshold to reduce false negative results.

We also report the results of our proposed network but without the separation loss (Eqn. 4.6) during training. This experiment was designed to highlight the importance of the separation loss in improving the accuracy of the predicted fiber orientations. We report the performance of all the algorithms on three synthetic tongue HARDI data sets (Section 4.3.2) and one post-mortem human tongue data set (Section 4.3.3).

Synthetic Tongue Data

In order to quantitatively evaluate the proposed fiber orientation reconstruction network, we simulated three single-shell HARDI data sets with the diffusion weighting of b=1000 s/mm^2 and 60 diffusion directions. The ground truth of the fiber orientations were generated based on a mechanical tongue model [123] (see Figure 4-2(b)). In total, there were 10,888 voxels containing fibers. In each voxel, there were at most two crossing fibers. The HARDI data were synthesized using a multi-tensor model [127]. Each fiber's diffusion tensor was computed by rotating a default single tensor with the desired fiber orientation. The default single-fiber tensor was set with diffusivities in orthogonal directions as $\{\lambda_1, \lambda_2, \lambda_3\} = \{0.002, 0.001, 0.001\}$ s/mm^2. The diffusivities were determined by fitting a diffusion tensor model on real HARDI tongue data. Although the diffusivity of the tongue muscle tissue is different from that of the brain tissue, the

SH coefficients of the tongue and brain HARDI data, which are the input to the proposed fiber orientation reconstruction network, have similar distributions. We thus believe that training this network with brain data is a reasonable surrogate, in the absence of tongue data. In voxels with crossing fibers, the fraction of the first fiber f_1 was sampled from a Gaussian distribution with mean of 0.6 and standard deviation of 0.05, and f_1 was set to 1 if the sample exceeded 1. The fraction of the second fiber f_2 is calculated by $1 - f_1$. We added three different levels of Rician noise to the synthesized signal; corresponding to SNR in the synthetic DWI of 10, 15, and 20.

Evaluation Metrics

The performance of the different methods were compared using the following criteria: angular correlation coefficient (ACC) [145], success rate (SR) [146], and angular error (AE) of the predicted fiber orientations. Detailed implementation of each evaluation metric is described below.

- **Angular correlation coefficient (ACC)**. ACC is a measurement that can be used to evaluate the similarity between two angular functions. In our experiments, the ACC, denoted as r_A, between the predicted and truth SH coefficients of the fODF was used to evaluate the performance of the fODF estimation. Given that the HARDI signal is real and symmetric and only even orders of SHs are used to fit the signal, the ACC is calculated as follows:

$$r_A = \frac{\sum_j c_j \hat{c}_j}{[\sum_j c_j^2]^{1/2} [\sum_j \hat{c}_j^2]^{1/2}} \tag{4.11}$$

  where c_j and $\hat{c}_j$ are the truth and predicted SH coefficients of the fODF, respectively. The total number of c_js is 45 in our experiments. The ACC values range from -1 to 1, where 1 indicates a perfect match.

- **Success rate (SR)**. To assess the correctness of the fiber number estimation, we computed the success rate (SR), which is defined as the proportion of voxels where the reconstruction algorithm correctly estimates the number of fiber orientations within a small angular tolerance. In this study, the tolerance was set to 20°, as suggested in [146]. To quantify the incorrect assessment, we also computed the false positive rate (FP), the false negative rate (FN), and the Dice score for the crossing fiber regions, which is an overlap measurement between the crossing fiber regions in the ground truth and the regions where the automatic algorithm predicted two fiber orientations.

- **Angular error (AE)**. We used AE between the estimated and true orientations to evaluate the angular accuracy of the estimated orientations. AE is defined as the angular difference between the true and predicted fiber orientations. Lower AE means more accurate fiber orientation reconstruction. AE was calculated only in the voxels where the reconstruction algorithm correctly predicted the number of muscle fiber orientations. In voxels with crossing fibers, the two estimated fibers were first matched to the ground-truth to minimize the error.

4.3.2   Fiber Orientation Reconstruction on Synthetic Data

In this experiment, we evaluated the performance of the proposed fiber orientation reconstruction network and the comparison methods on three synthetic tongue data sets. There were 10,888 voxels containing fibers in total. Figure 4-7 presents the boxplot of the ACC between the true and predicted SH coefficients produced by CSD and the proposed CNN method without or with the separation loss. ACC was not computed for BEDPOSTX since it does not output the fODF. We see that both the proposed CNN methods (without and with the separation loss) provided higher ACC than the CSD method, especially when the SNR is low.
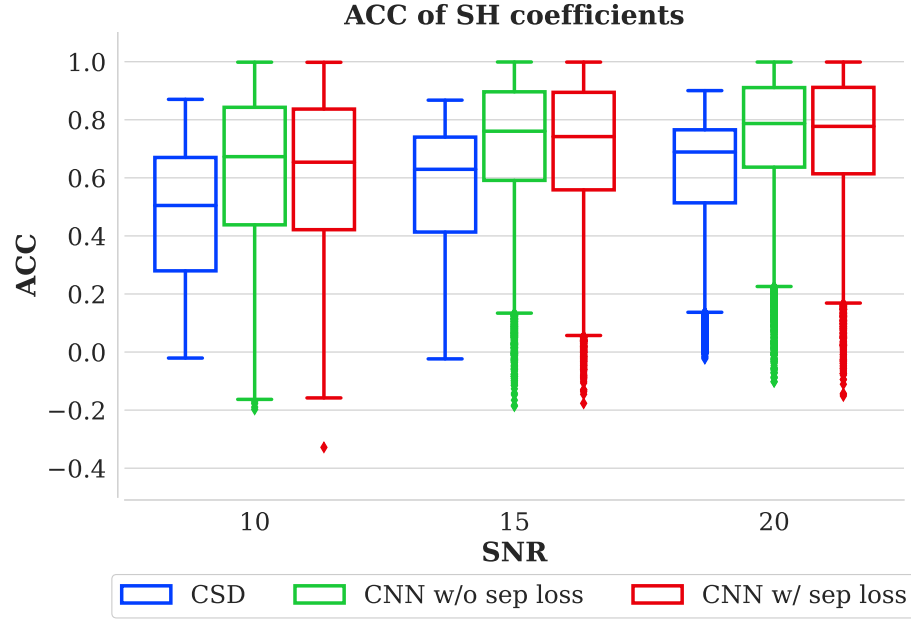
Figure 4-7. Boxplots of the angular correlation coefficients (ACC) for the spherical harmonic (SH) coefficients of the fiber orientation distribution function (fODF) produced by CSD and the proposed fiber orientation reconstruction network without or with the separation loss. ACC was not computed for BEDPOSTX since it does not output the fODF.

Table 4-I presents the performance of each method on the reconstruction of the number of fiber orientations. The proposed CNN method with the separation loss outperforms all the other methods at each noise level in terms of SR and FN. The overestimated metric FP of the proposed method is higher than the best method, but remains at a low level. The mean Dice scores on the crossing fiber regions for each method are 0.57, 0.23, 0.38, and 0.64, respectively. The proposed method shows better accuracy estimating the number of fibers in crossing fiber regions.

Figure 4-8 presents the boxplots of the AE in single (left) and crossing (right) fiber orientation regions. We can observe that in single fiber regions, the CNN methods produced slightly more accurate fiber orientation when the noise level is high, which is closer to real tongue DWI. In crossing fiber regions, the proposed CNN with the separation loss shows the lowest AE compared to all the other methods. It is worth

Table 4-I. Success rate (SR), false positive rate (FP), and false negative rate (FN) on three synthetic HARDI tongue data with different noise levels. The results were produced by the following fiber reconstruction algorithms: CSD, BEDPOSTX, proposed CNN without the separation loss, and proposed CNN with the separation loss. The total number of voxels is 10,888. Bold: best performance among the four reconstruction algorithms.

SNR=10			
	SR	FP	FN
CSD	0.451	0.036	0.513
BEDPOSTX	0.328	0.016	0.656
CNN w/o l_{sep}	0.439	**0.006**	0.555
CNN w/ l_{sep}	**0.546**	0.012	**0.442**

SNR=15			
	SR	FP	FN
CSD	0.616	0.007	0.376
BEDPOSTX	0.389	0.020	0.591
CNN w/o l_{sep}	0.495	**0.003**	0.502
CNN w/ l_{sep}	**0.674**	0.007	**0.319**

SNR=20			
	SR	FP	FN
CSD	0.715	**0.001**	0.284
BEDPOSTX	0.517	0.028	0.455
CNN w/o l_{sep}	0.534	0.002	0.464
CNN w/ l_{sep}	**0.717**	0.006	**0.277**

noting that when removing the separation loss, the same network architecture did not show much improvement compared to CSD and BEDPOSTX in crossing fiber regions. This clearly demonstrates that the separation loss contributes to improving the accuracy of the predicted fiber orientations.

Figure 4-9 shows the SR and the mean AE simultaneously on single (circles) and crossing (triangles) fiber orientation regions. In each plot, the upper left point is the optimal point. In the single fiber voxels, the proposed CNN method with the separation
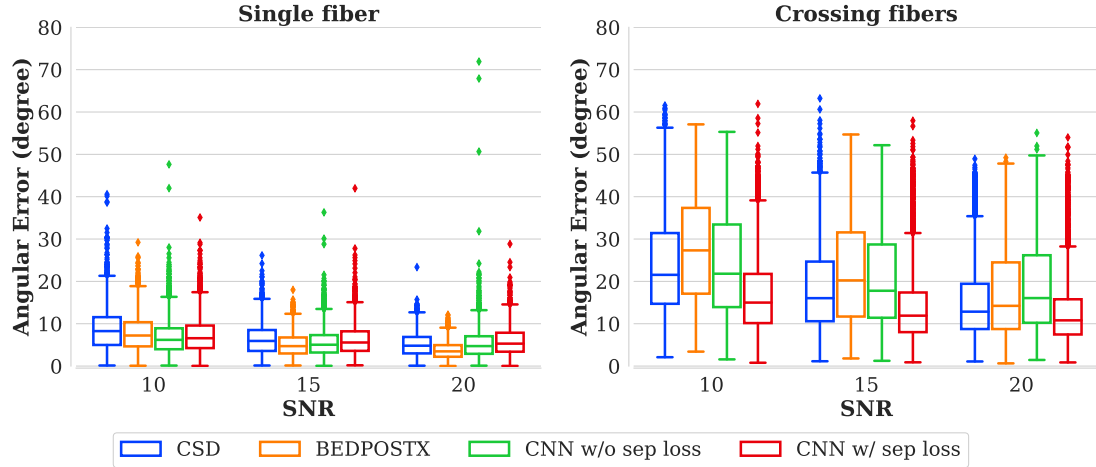
Figure 4-8. Boxplots of the angular error in voxels with single (left) and crossing (right) fiber orientations.

loss (red circles) shows competitive results compared with other methods in each noise level. In regions with crossing fiber orientations, the proposed method (red triangles) shows better performance than the other methods.

4.3.3   Fiber Orientation Reconstruction on Real Data

We applied the fiber orientation reconstruction methods to a post-mortem human tongue data and show qualitative comparison. The post-mortem human tongue data were obtained on an 86-year-old female, around 48 hours post death. Use of post-mortem specimen was approved by the Institutional Biosafety Committee. The diffusion MRI was acquired on a Siemens Prisma scanner using a Readout-Segmented Echo-Planar diffusion sequence with the following parameters: readout segment size = 3, TE = 70 ms, TR = 6230 ms, in-plane resolution = 2.5 mm, slice thickness = 2.5 mm, sampling bandwidth = 1645 Hz/Px, and echo spacing = 0.34 ms. There were 14 non-diffusion weighted images (b0 images) and 200 diffusion gradient directions with b-value of 2000 s/mm^2. A tongue mask was manually drawn to restrict the processing region.

A visual comparison of the fiber orientation reconstruction results produced by the
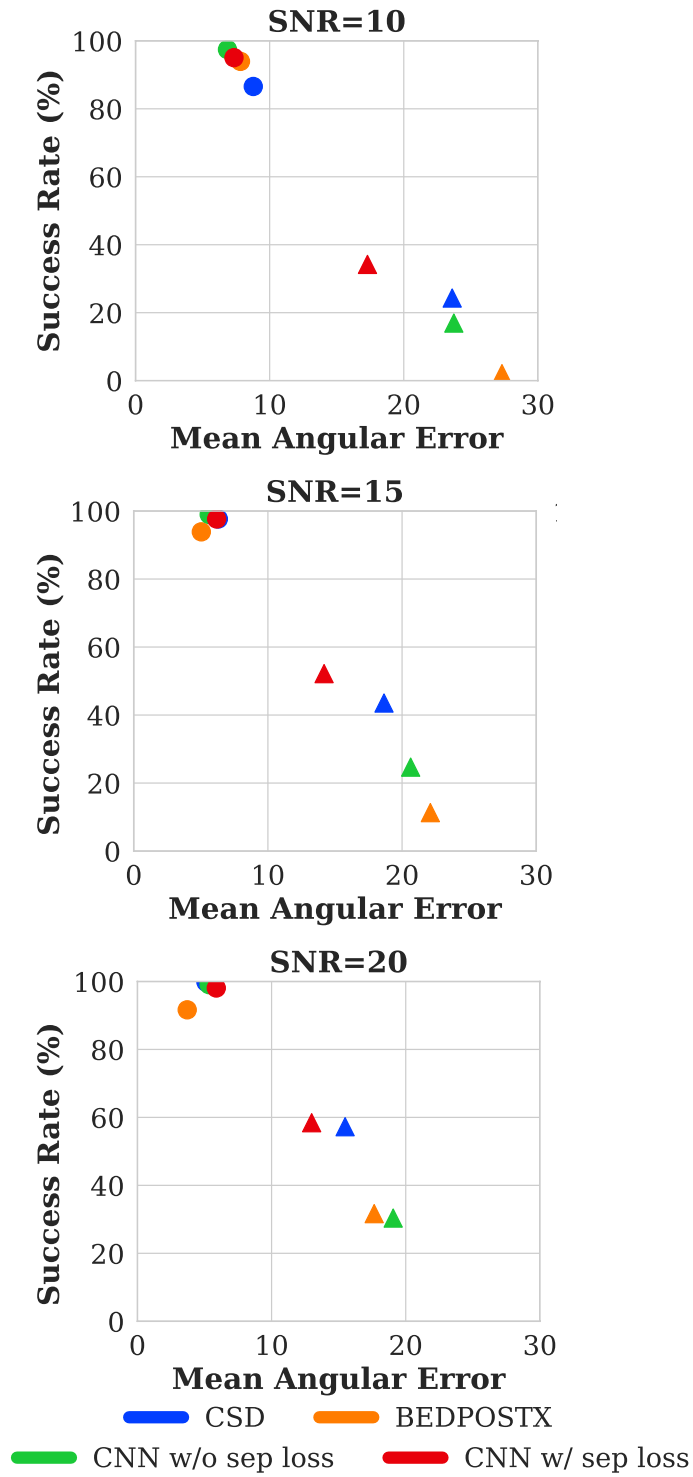
Figure 4-9. Success rate vs mean angular error in single (circle) and crossing (triangle) fiber regions.
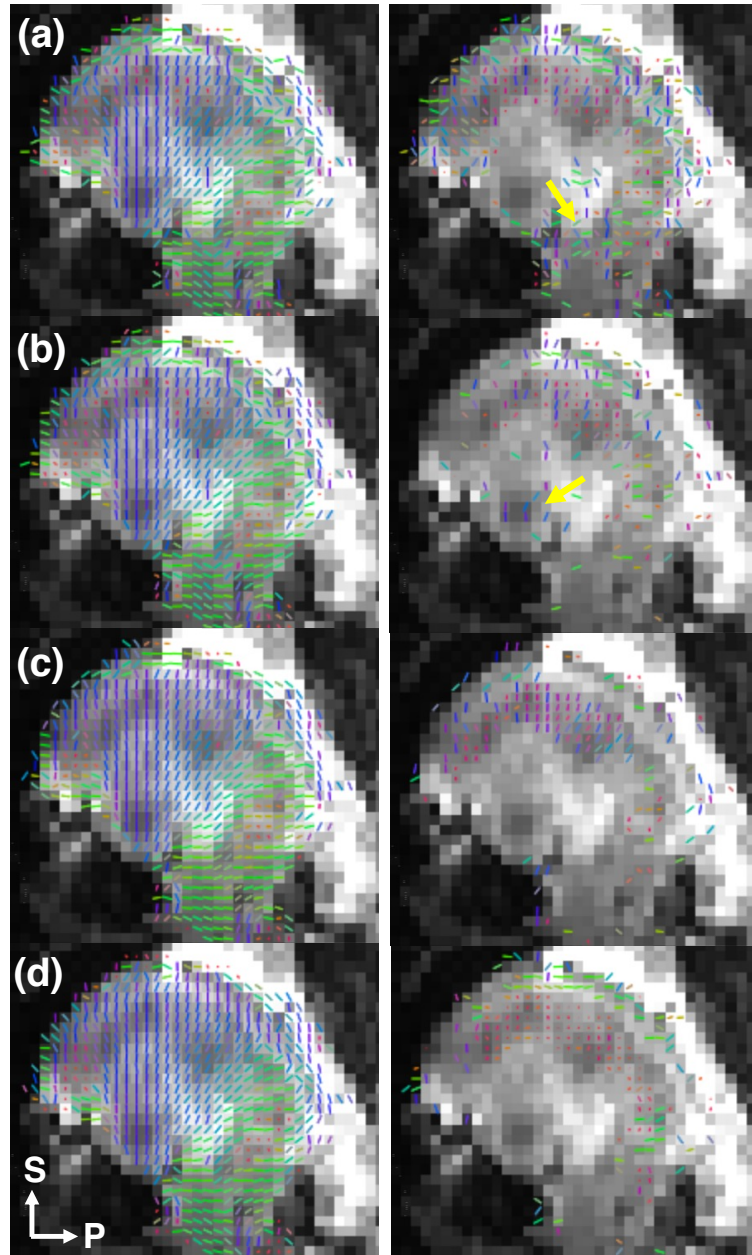
Figure 4-10. Sagittal view of the fiber orientations produced by the four fiber orientation reconstruction methods on a post-mortem human tongue data: (a) CSD; (b) BEDPOSTX; (c) CNN without the separation loss; (d) CNN with the separation loss. S: superior; P: posterior. Left column shows the first orientation and right column shows the second orientation. The yellow arrows in (a) and (b) point to some false positive prediction on the second orientation. The fiber orientations are conventionally color-coded (red: right-left, green: anterior-posterior, blue: inferior-superior).

four methods is shown in Figure 4-10. The predicted fiber orientations are overlaid on a b0 image. We observe that all the methods provided similar results on the first fiber orientation. However, CSD and BEDPOSTX provided more false positive predictions on the second fiber orientation in areas where only GG muscle should exist (see the yellow arrows in Figure 4-10(a)(b)). Compared to the other three methods, our CNN with the separation loss provides cleaner and more consistent fiber predictions. Comparing Figure 4-10(c) and (d) reveals the effect of the separation loss. In regions where GG and T muscles are crossing, our CNN with the separation loss produced more accurate T muscle orientations, which are orthogonal to GG muscle and should be perpendicular to the sagittal plane. We can also observe that some fiber assignments are inconsistent. In some regions where T and V, or T and GH muscles are crossing, the T muscle orientation appeared as the first orientation instead of the second one. The reason could be that the strength of the T muscle is higher in these regions. The inconsistent assignments can be corrected by the proposed fiber matching algorithm (described in Section 4.2.5 and 4.2.6.

4.3.4   Experimental Setup for Fiber Matching

In order to quantitatively evaluate the proposed fiber matching algorithm, we simulated 72 pairs of fiber orientation images with part of the voxels having incorrect assignments. The original ground truth fiber orientations were generated based on a mechanical tongue model [123] (see Figure 4-2(b)). There were 10,888 voxels containing fibers and each voxel consists of at most two crossing fibers. The testing images were created by two steps: the first step is to rotate the fibers in each voxel along a random axis by a small random degree, and the second step is to randomly select some voxels and switch the fiber assignments to simulate the output from the fiber reconstruction method. The degree of random rotation follows a uniform distribution $\mathcal{U}(-\theta, \theta)$, and θ was set to

0°, 10°, 20°, and 30° in our experiments. The ratio of switched voxels r was set to the following levels: 0.1, 0.2, ..., and 0.9. For each combination of θ and r, 2 pairs of fiber orientation images were generated with independent realizations.

4.3.5 Fiber Orientation Matching Results

The performance of the proposed method was evaluated using the assignment error rate, which is defined as the number of voxels with incorrect assignment divided by the total number of voxels that contain fibers (10,888 in our case). Table 4-II presents the assignment error rate after applying the fiber matching algorithm on the synthetic fiber images. The proposed method greatly reduced the number of voxels that have incorrect assignments.

Table 4-II. Mean assignment error rate (%) after fiber matching on synthetic fiber orientation images. Each combination has two pairs of fiber images. The assignment error rate before correction is: $\min\left(r, 1-r\right) \times 100\%$

θ \ r	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0°	0.101	0.101	0.101	0.101	0.101	0.101	0.101	0.101	0.101
10°	0.124	0.119	0.106	0.106	0.101	0.092	0.106	0.092	0.142
20°	0.161	0.161	0.152	0.133	0.184	0.151	0.156	0.142	0.147
30°	0.193	0.303	0.225	0.248	0.202	0.202	0.230	0.216	0.243

It is worth noting that the objective function (Equation 4.8) does not enforce the first orientation to be the first after correction. Our goal is to create consistent fiber orientation images. Therefore, when $r > 0.5$, the algorithm is likely to generate completely inverse first and second orientation images. Figure 4-11 shows two visual results of the fiber matching on the synthetic fiber images with $r = 0.4, \theta = 30°$ and $r = 0.7, \theta = 20°$. In the mid-sagittal slice of the case where $r = 0.4, \theta = 30°$, the first fiber orientation image contains the GG muscle. And the second fiber orientation image contains the SL and
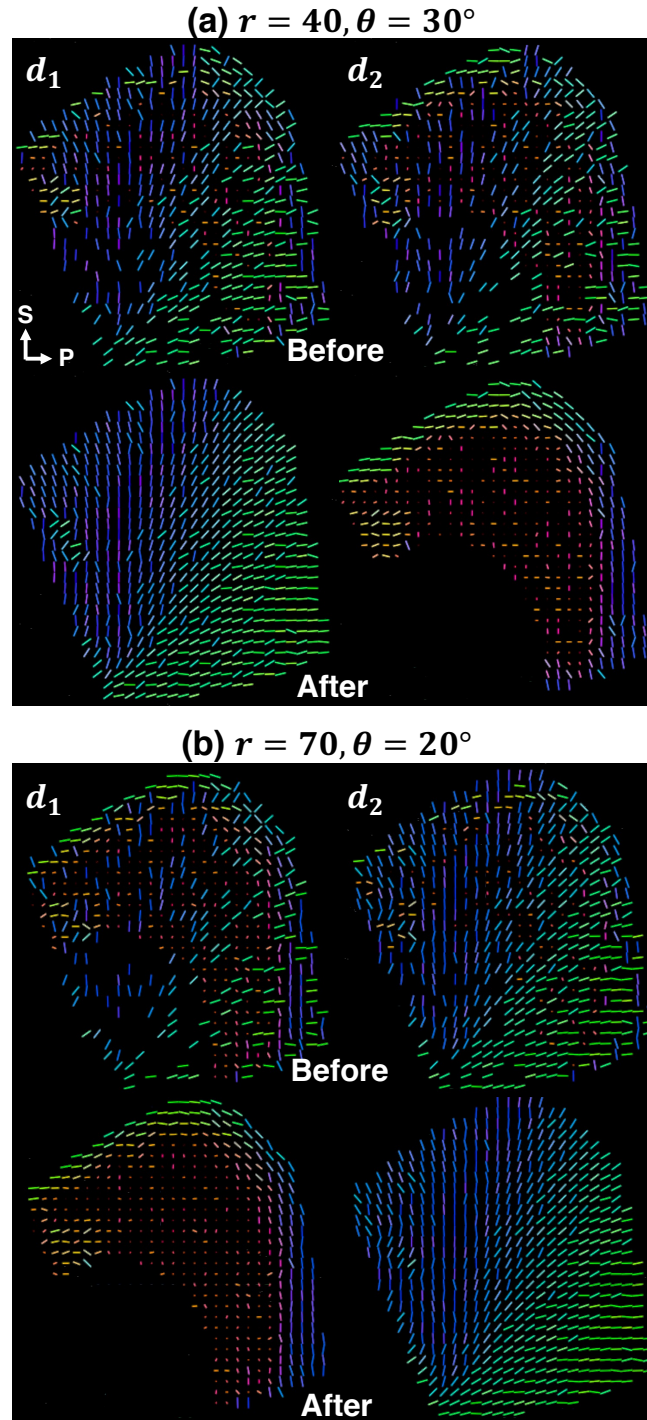
Figure 4-11. Mid-sagittal slice of the fiber matching results on synthetic tongue muscle fibers with random rotation and switched fiber orientations. (a): Before and after fiber matching with $r = 0.4, \theta = 30°$. (b): Before and after fiber matching with $r = 0.7, \theta = 20°$. The fiber orientations are conventionally color-coded (red: right-left, green: anterior-posterior, blue: inferior-superior).

T muscles. In the case where $r = 0.7, \theta = 20°$, the algorithm produced completely opposite first and second fiber images. In both cases, the fiber orientations are more consistent after correction.

We applied the fiber matching algorithm on data from a post-mortem human tongue. The data acquisition is described in Section 4.3.3. Figure 4-12 shows a visual comparison of the fiber orientation reconstruction results produced by the fiber orientation reconstruction network (in Section 4.2.2) and refined by the fiber matching algorithm. Figures 4-12(a) and (c) show that the fiber orientation reconstruction network can resolve crossing muscle fiber in the tongue and predict orthogonal fiber orientations. However, we can observe that in some regions where GG and T muscles are crossing (see yellow contours in Figure 4-12), the T muscle fibers appear as the first orientation, which is not consistent with most other regions of the T muscle. The reason for this inconsistency could be that the strength of the T muscle is higher than other muscles in those regions. From Figures 4-12(c) and (d), we see that the fiber matching algorithm switched the fiber assignments in those regions and provided cleaner fiber orientations.

4.4 Discussion

We have performed evaluation of our proposed fiber orientation reconstruction network and the fiber matching algorithm using both synthetic and real tongue data. Quantitative evaluation was done on synthetic data and qualitative results were shown on post-mortem tongue data. The accuracy of the estimated fiber orientations was evaluated using angular correlation coefficients, success rate, and angular error. When testing on the synthetic tongue data generated by a mechanical tongue model, our method produced more accurate assessment of the number of the tongue muscle fiber orientations and better fiber orientation reconstruction compared to widely used fiber reconstruction algorithms. As shown in Figure 4-9, our method achieved higher success

90
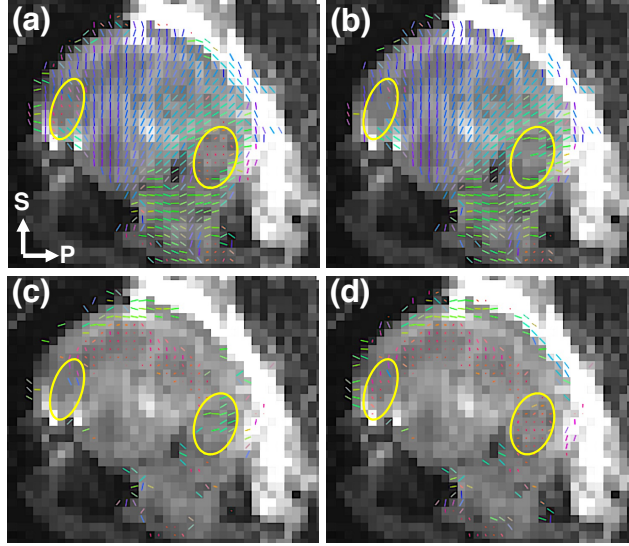
Figure 4-12. Sagittal view of the fiber orientations on data from a post-mortem human tongue. The underlying image is a b0 image. (a) and (c): first and second fiber orientations directly produced by the proposed fiber orientation reconstruction network; (b) and (d): first and second fiber orientations refined by the fiber matching algorithm. Yellow contours point to some regions where the fiber assignments were corrected by the fiber matching algorithm.

rate and lower angular errors. When evaluating on the real tongue data, the proposed method provided cleaner and more consistent fiber predictions.

Considering the fiber orientation reconstruction process, our proposed method can be run in a fully automatic manner. In contrast in CSD we must manually select the region of the non-crossing area to create the response function required by the algorithm. Some software packages [147] can produce the response function automatically, but the output of the CSD algorithm is only the fiber orientation distribution function. Given this, a threshold must be manually set to extract the peak fiber orientations. This is the same in BEDPOSTX, where a threshold is also required to remove non-dominant fibers.

In the experiments, we also compared the results with the same network architecture but without the separation loss. The same network can produce a better fiber orientation distribution function, but it produced worse fiber orientation reconstruction results, demonstrating the significance of the proposed loss function.

The fiber matching algorithm was proposed to correct inconsistent assignments of the fiber orientations produced by the fiber orientation reconstruction network. It was quantitatively evaluated on synthetic tongue fiber images with random rotations and switching. Furthermore, qualitative results on a post-mortem tongue demonstrated its ability to improve visualization of the tongue muscle fibers.

The proposed fiber matching algorithm can also be used as a post-processing step for other fiber reconstruction methods that produce one or two fiber orientations per voxel. Figure 4-13 shows the fiber predictions on the same post-mortem tongue produced by the CSD method, and the corresponding fiber matching results. We can observe that our method improves the visualization of the fiber orientations for further analysis.
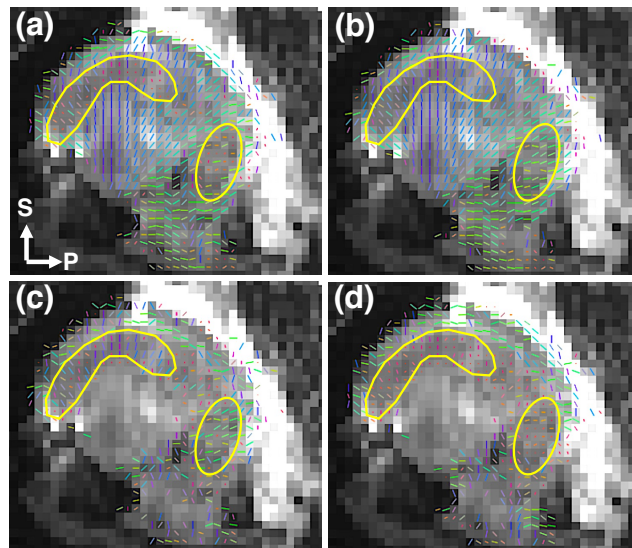
Figure 4-13. Sagittal view of the fiber orientations on data from a post-mortem human tongue. The underlying image is a b0 image. (a) and (c): first and second fiber orientations produced by CSD method; (b) and (d): first and second fiber orientations refined by the fiber matching algorithm. Yellow contours point to some regions where the fiber assignments were corrected by the fiber matching algorithm.

4.5 Summary

In this chapter, we presented a 3D patch-based convolutional neural network to recon-
struct the tongue muscle fiber orientations from HARDI data. We introduced a separation
loss that encourage the network to produce orthogonal crossing fiber orientations, which
is an important characteristic in the human tongue. We also presented a fiber matching
algorithm to refine the reconstructed fiber orientations and produce more consistent
results. The algorithm is implemented as an unconstrained quadratic binary optimization
model which maximizes the similarity between the first or second fibers in neighboring
voxels.

We showed that our proposed methods can produce more accurate and consistent
fiber orientations than competing methods on synthetic and real tongue data. Recon-
struction of the interdigitated tongue muscle fiber orientations provides the possibility to
characterize and analyze the interaction and cooperation between the muscle groups
during tongue deformation, which we will introduce in Chapter 5.

Chapter 5

Tongue Strain Analysis Pipeline

5.1 Introduction

5.1.1 Background

The human tongue plays an essential role in many vital behaviors including swallowing, breathing, and speech production [119, 148]. In speech production, the tongue helps to form sounds by changing its surface shape via highly complex deformation patterns [149, 150] that are produced by sequential activation of different tongue muscles. Understanding and interpreting the relationship between individual muscle groups and the complex functions performed by the tongue have been previously investigated [120, 151, 152], but this research has been hampered by the inability to directly measure muscle activations in the tongue. Electromyography (EMG) has been used to record the electrical activity of tongue muscle during rest and contraction for diagnostic purposes [153–155]. The EMG signals are measured by placing electrodes over muscle groups of interest to record the electrical activity. However, the surface or needle electrodes that are used to make these measurements affect natural speech motion.

Measurements of three-dimensional strain throughout the tongue using MRI have

been used as proxy measurements through which to infer muscle activity [151, 156]. Measurements made this way do not interfere with natural speech production (other than the speaker is lying in the supine position), but previous inferences about the relationships between strain and specific muscle groups have been limited by absent or inaccurate identification of tongue muscle fibers within the tongue. Such identification is challenging because the tongue contains a collection of extensively interdigitated muscles that are nearly orthogonal in three dimensions [119, 122]. We have explored the tongue muscle fiber reconstruction in Chapter 4.

In the study of speech production, analyzing the interaction and cooperation between tongue muscles is of great scientific and clinical interest [32, 33]. Also, understanding the adaptive behaviors of the tongue in patients with prior glossectomies—e.g., where tongue has been partly surgical removed due to tongue cancer—has been of interest to oral surgeons and speech pathologists [157–160]. In this chapter, the overall goal is to analyze and compare muscle cooperation pattern differences between the healthy controls and glossectomy patients in speech production. We obtained insights related to this using imaging methods that permit characterization of muscle fiber orientations, muscle group delineations, and muscle strain analysis.

5.1.2 Tagged and Cine MRI

MRI is able to reveal both anatomical structures and tissue motion. Estimation of tongue motion has previously been performed using tagged MRI [34, 152, 161]. In tagged MRI acquisition, a temporary pattern is made to appear in the images using manipulation of the magnetic field [162]. As the tongue moves during speech production, the tag patterns can deform with tissue during motion. The internal motion of the tongue can be reconstructed by processing a sequence of tagged MRI acquired over time [34, 163]. Figures 5-1(a) and (b) show examples of horizontal and vertical tagged MR slices,
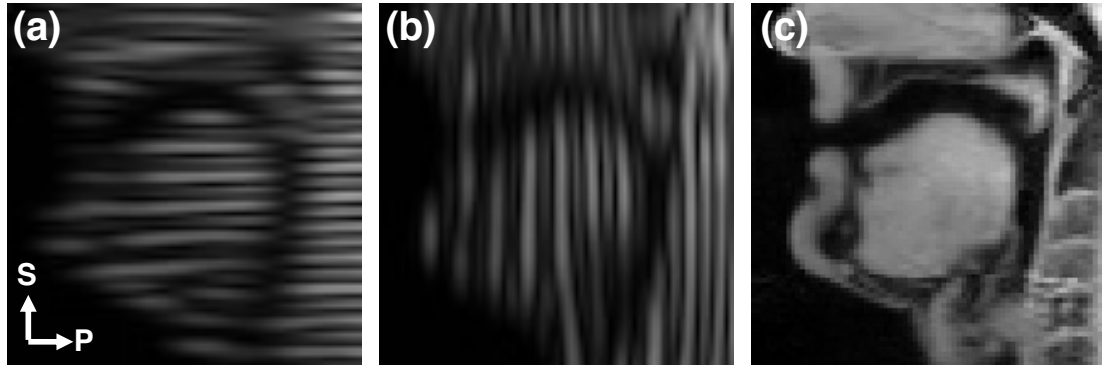
Figure 5-1. (a) Mid-sagittal slice of a horizontal tagged MRI of a human tongue. (b) Mid-sagittal slice of a vertical tagged MRI of a human tongue. (c) Mid-sagittal slice of a cine MRI of a human tongue.

respectively.

Although the tagged MRI can capture the internal motion of the tongue, it has low spatial resolution and provides low anatomical information at the tissue-air boundary. Therefore, we collected an additional sequence of MR images without tags, called cine MR images, to characterize the boundary of the tongue tissue. Figure 5-1(c) shows an example of a cine MR image from a human tongue.

5.1.3 Related Work

Previous studies have used tagged MRI to compare the motion fields produced by speech in healthy controls and post-partial glossectomy patients [3, 4]. To study the interaction of tongue muscles, Xing et al. [156] analyzed strain in tongue muscle fibers and showed contraction-extension patterns of individual muscles during speech. Xing et al. [152] identified the muscle coordination patterns among various tongue muscles in a normalized space to investigate speech motor control. Gomez et al. [151] presented a model-based approach to study the mechanical cooperation of the tongue muscles.

To analyze the muscle cooperation pattern in speech production, both accurate

reconstruction of the muscle fibers and the estimation of the tongue motion are required. Although use of both diffusion MRI and tagged MRI in the tongue is possible, there is very little work on their joint use. Existing works that do consider both together [156, 164], do not characterize different strains in the orthogonal fibers that interdigitate within the tongue, which constitutes about two-thirds of the entire tongue. Also, the strains that characterize specific muscles are not properly distinguished since their interdigitated fiber groups are not separately associated with their underlying tongue muscles. Without these capabilities, it has not been previously possible to characterize contractions and elongations in whole tongue muscles during speech and, accordingly, it has not been possible to characterize the coordinated interactions between tongue muscles in their generation of tongue deformations for speech production.

5.1.4 Contribution

In this chapter, we present a comprehensive pipeline that jointly analyzes diffusion MRI, tagged MRI, and cine MRI to reveal strain components in muscle fibers throughout the tongue during speech production. We present a measurement and analysis process that estimates fiber directions throughout the tongue—including regions of crossing fibers—and determines strain components along all these computed fiber directions. This permits a far more detailed analysis of tongue muscle strains throughout the tongue during speech production. The proposed pipeline incorporates a set of algorithms. The tongue muscle fiber orientation reconstruction network and the fiber matching algorithm described in Chapter 4 are used to produce consistent muscle fiber orientations from diffusion MRI. A phase vector incompressible registration algorithm [34] is used to estimate the tongue motion during speech from tagged MRI. Given these measurements, the motion fields were temporally and spatially aligned across all subjects before further analyses, and strains in the line of action (SLAs) at each tongue voxel and for each

fiber direction were calculated by projecting the computed strains onto the local fiber directions. Finally, using a manually labeled high-resolution muscle mask, the SLA patterns in two individual muscle groups were analyzed to reveal the cooperation between these muscles. We performed these steps and this analysis on a cohort of 8 healthy controls and 2 post-partial-glossectomy patients and achieved both qualitative and quantitative results. This new approach has the potential to aid researchers in better understanding speech production and in learning about the adaptive behaviors of patients with glossectomies.

5.2   Methods

5.2.1   Data Collection

The data set in this study includes 8 healthy controls and 2 post-partial-glossectomy patients with tongue flaps. All images were acquired on a 3-T Prisma MR scanner (Siemens Healthcare), with a 64-channel head/neck coil. To reconstruct the tongue muscle fiber orientations, we collected HARDI data for each subject in a static state. The HARDI tongue data were acquired using a single-shot echo-planar diffusion MRI sequence with real-time motion detection and re-acquisition [165, 166] with TE = 48 ms and TR = 2600 ms. The image resolution is 2.5 mm isotropic and 34 axial slices were acquired to cover the entire tongue. There were 14 non-diffusion weighted images (b0 images) and 200 diffusion gradient directions with b-value of 500 s/mm^2.

To capture the tongue's motion during speech, the participants were asked to say the phrase "a thing" in repeated speech cycles while tagged and cine MR images were acquired. The phrase was designed to start with a neutral tongue position /ə/, moving the tongue tip forward to make /θ/, and ending with the upward motion of the tongue body into /ŋ/. The motion into the /θ/ uses the most tongue tip protrusion of

any English sound. As described in Section 5.1.2, although tagged MRI provides information about internal tongue deformation, its images have low spatial resolution. Therefore, we use high resolution MRI and cine MRI to characterize tongue anatomy. Both tagged and cine data were collected over multiple repetitions of the speech task, timed to a metronome repeated every 2 seconds. The duration of the recorded phrase is 1 second, and 26 time frames (TFs) were acquired over this period. The in-plane resolution is 1.875×1.875 mm and the slice thickness is 6 mm. The tagged MRI data were collected using a CSPAMM pulse sequence [167] in both sagittal and axial orientations. In the sagittal acquisitions, vertical and horizontal tags were used to capture the anterior-posterior and superior-inferior motion, respectively. The left-right motion was captured using vertical tags in the axial acquisitions. The cine MRI data were collected in axial, coronal, and sagittal orientations, and these acquisitions were combined using a super-resolution algorithm [168] to form a single high-resolution image volume.

5.2.2 Flowchart of the Tongue Strain Analysis Pipeline

The overall processing pipeline is illustrated in Figure 5-2. The input to the pipeline is HARDI data as well as tagged and cine MRI of each subject. The output is the strain line the line of actions (SLAs) associated with the muscle fiber orientations. The details of each part in the pipeline are described in the following sections.

5.2.3 Tongue Muscle Fiber Orientation Reconstruction

To reconstruct the tongue muscle fiber orientations, we applied our fiber orientation reconstruction network and the fiber matching algorithm to HARDI tongue data, as described in Chapter 4. For each subject, a tongue mask was manually delineated to restrict the processing region. The HARDI tongue data was first transformed to SH coefficients and sent to the fiber orientation reconstruction network. Then the two fiber

99
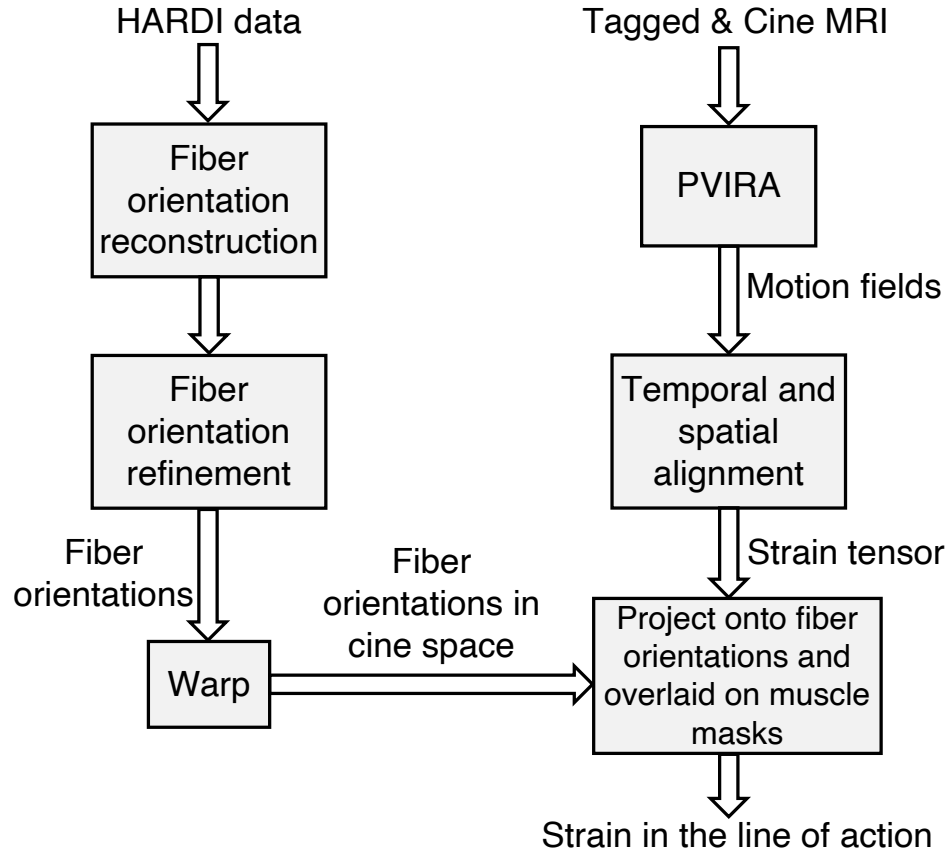
Figure 5-2. Flowchart of the tongue strain analysis pipeline. HARDI = high angular resolution diffusion imaging.

orientation images produced by the network were input to the fiber matching algorithm to create clean and consistent fiber orientations. In this project, the goal is to analyze the strain along the local muscle fiber orientations. Therefore, it is important to correct the assignments of the fiber orientations for the purpose of computing average strains within specific muscle groups.

5.2.4   Tongue Motion Estimation and Alignment

To estimate tongue motion during speech, tagged MRIs were processed by a phase vector incompressible registration algorithm (PVIRA) [34]. For each subject, the result is

a dense 3D incompressible motion field at each time frame, and there are 26 motion fields in total for the phrase. In this work, we estimated the motion field frame by frame, which means the reference state for each time frame is its previous time frame. Then the motion field from any time frame to another time frame can be computed by composing the multiple motion fields.

To yield a consistent analysis across subjects, we must align the subjects' motion fields in both space and time. Temporal alignment is important because subjects speak at different rates. There are four critical time frames, /ə/, /θ/, /i/, and /ŋ/ in the phrase "a thing". These times are determined for each subject by manual inspection of the mid-sagittal slices of the cine MRI throughout the image sequence. We then follow the instructions in Xing et al. [152] to align the subjects temporally. Briefly, we first reassign the four critical time frames to a set of pre-defined critical time indices $t_ə$, $t_θ$, t_i, and $t_ŋ$, which are 1, 8, 14, 20 in our experiments. We then interpolate the remaining time frames between these critical ones by linear interpolation using the two closest time frames in the subject's original data. This yields a set of time-aligned volumetric images (both cine and tagged MRI) of each subject saying the phrase "a thing".

To spatially align the motion fields of our subjects, we construct a cine atlas at time index $t_ə$ and deform the motion fields from different subjects to the atlas. The cine atlas is constructed by groupwise registration among the cine MR images at $t_ə$ from all the healthy controls. Then we register the cine MR images at $t_ə$ of each subject, including the patients, to the cine atlas using diffeomorphic deformable image registration [51]. The deformation field between each subject and the cine atlas is referred to as ϕ. After that, each subject's PVIRA motion fields can be deformed to the cine atlas space by composing ϕ and a series of PVIRA motion fields in the original subject's space [152]. After all the subjects' motion fields are temporally and spatially aligned, we can compute the strain tensor for further SLA analysis.

5.2.5   Tongue Muscle Mask Generation

Analysis of the averaged SLAs in different muscle groups enables us to study how particular muscles interact and collaborate over time during speech production. A manually-labeled tongue muscle mask in a T2-weighted (T2-w) high-resolution MRI atlas [169] was used to segment the tongue muscles of the healthy subjects. Basically, the first step is to find a deformation field between the T2-w atlas and each subject's cine MR image at t_∂ by rigid and deformable registration between the tongue masks in the two images. Next, the manually-labeled muscle mask in the T2-w atlas space is transformed by this deformation field to generate muscle masks for each subject. The rationale for using tongue masks instead of the intensity images to perform the registration is that the T2-w MRI was acquired in a static state while the tagged/cine data were acquired during speech production. The air gap between the top of the tongue and palate in the tagged/cine data makes registration to the static-state intensity images inaccurate. For the patient data, individual muscle masks, including the flap regions, were manually delineated in the patient's T2-w MR image and were deformed to the cine space using the same registration method as in the healthy subjects. The ratio of the flap volume to the whole tongue volume is 0.17 for patient 1 and 0.26 for patient 2.

We focus on one extrinsic muscle, the genioglossus (GG) muscle, and one intrinsic muscle, the transverse (T) muscle to demonstrate the pipeline. Detailed description of the tongue muscle groups can be seen in Section 4.1.1. These two muscles were further subdivided into anterior (GGa, Ta), middle (GGm, Tm), and posterior (GGp, Tp) parts. Figure 5-3 shows the manual labeling of the GG and T muscle groups on the T2-w MRI atlas (Figure 5-3(a)) and the two glossectomy patients (Figures 5-3(b) and 5-3(c)). The flap region is also shown in the patient data.
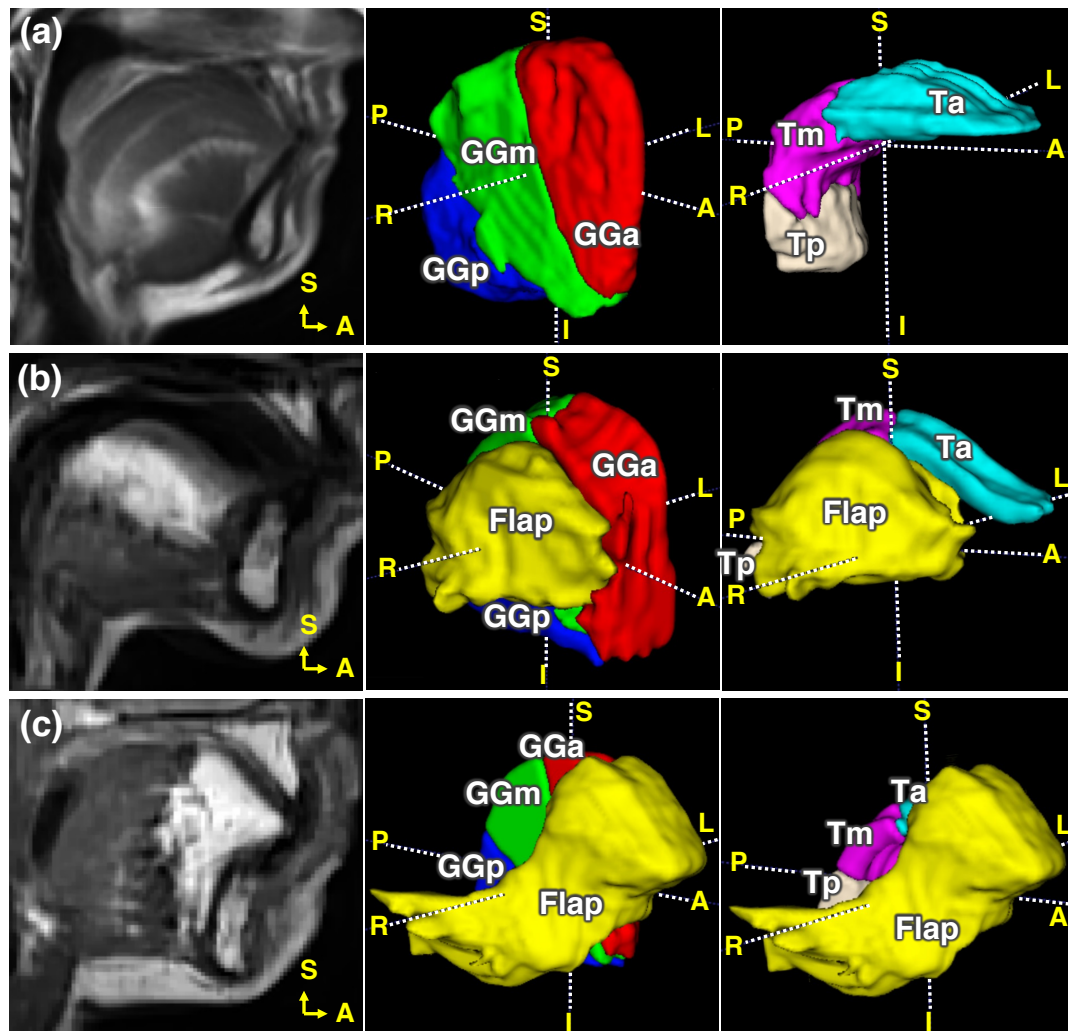
Figure 5-3. High-resolution T2-w MR images and the manual labeling of the genioglossus (GG) and transverse (T) muscles using the T2-w images. The two muscles are subdivided into anterior (a), middle (m), and posterior (p) parts. In each row, the left subfigure shows a sagittal slice of the T2-w MR image. The middle and right subfigures show 3D rendering of the GG and T muscles, respectively. Flap region is also shown in the patients. (a): Manual labeling of the muscles on the T2-w MR image atlas. (b) and (c): Manual labeling of the muscles and flap on the two glossectomy patients P1 and P2, respectively. S = superior; I = inferior; A = anterior; P = posterior; R = right; L = left.

5.2.6 Strain in the Line of Action

Given the previous steps, the strain tensor can be now computed from the aligned motion fields and projected onto the local fiber orientations to calculate the strain in the line of action (SLA). The SLA is defined as a ratio of a local tissue line element that deforms along the muscle fiber direction compared to its length in the reference time frame, and is calculated as

$$\lambda(\boldsymbol{d}_m(i)) = \sqrt{\boldsymbol{d}_m(i)\boldsymbol{C}_{ij}\boldsymbol{d}_m(i)}, m \in \{1, 2\} \tag{5.1}$$

where $\boldsymbol{C}_{ij}$ is the right Cauchy-Green deformation tensor and is defined as $\boldsymbol{C}_{ij} = \boldsymbol{F}^T\boldsymbol{F}$, and $\boldsymbol{d}_m(i)$ is the first or second muscle fiber orientation at the ith key time frame. The $\boldsymbol{F}$ in the definition of $\boldsymbol{C}_{ij}$ is the deformation gradient tensor from ith key time frame to the jth one. It is a 3×3 matrix in 3D coordinate system and is defined as

$$\boldsymbol{F} = \frac{d\boldsymbol{x}_j}{d\boldsymbol{X}_i} \tag{5.2}$$

where $\boldsymbol{x}_j$ is the coordinate of a tissue point in the deformed time frame j and $\boldsymbol{X}_i$ is the coordinate of the same tissue point in the reference time frame i. SLA values greater than 1 represent extension and values less than 1 represent contraction; for example, when SLA has the value 1.1 this can be interpreted as a 10% extension while an SLA value of 0.9 can be interpreted as a 10% contraction. In the experiments, we apply the Lagrangian framework and all the motion fields are measured in the reference time frame, which is t_ϑ in the experiments. Another choice is the Eulerian framework, where all the motion fields are measured in the deformed time frame. Since the HARDI data were acquired in static state, it is better to map and display the SLA quantity computed at each time frame to a reference time frame.

Since all the motion fields have been spatially aligned to a cine atlas space (Sec-

tion 5.2.4), we need to map the muscle fiber orientations, $\boldsymbol{d}_m, m \in \{1, 2\}$, at each voxel to the same cine atlas space to calculate the SLA. For each subject, we first find the transformation between the tongue masks in the HARDI b0 image and the cine atlas by applying rigid and diffeomorphic image registration. Then the muscle fiber orientations, $\boldsymbol{d}_m, m \in \{1, 2\}$ estimated on the HARDI data are deformed into the cine atlas space by this transformation. The reason for using tongue masks instead of MR images is the same as in the description in Section 5.2.5.

5.3 Experiments and Results

5.3.1 Fiber Orientation Estimation

The fiber reconstruction network and the fiber matching algorithm were applied to the HARDI data of all 10 subjects. The 8 healthy controls are referred to as C1, C2, ..., C8, and the 2 patients are referred to as P1 and P2. A whole tongue mask was manually drawn for each to restrict the processing region. Figure 5-4 shows a visual comparison of the predicted fiber orientations, overlayed on a b0 image, of a healthy control (C8, shown in Figure 5-4(a)) and a patient (P2, shown in Figure 5-4(b)). In each sub-figure, the upper row shows the first fiber orientation and the lower row shows the second fiber orientation. In the mid-sagittal slice (Figure 5-4(a) top), we can observe that in the healthy control the GG fiber orientation, as expected, is fan-shaped and covers the whole body of the tongue (blue and green lines). In Figure 5-4(a) bottom, the T muscle, represented by red dots, is arc-shaped, following the upper surface shape of the tongue, and left-right in orientation. Note that the algorithm does not calculate a second fiber orientation in the region of the tongue that is composed solely of GG fibers. The fiber orientations of Ta (bottom) intersect with the GG fiber orientations (top), at approximately right angles. In the patient data (Figure 5-4(b) top and bottom), the
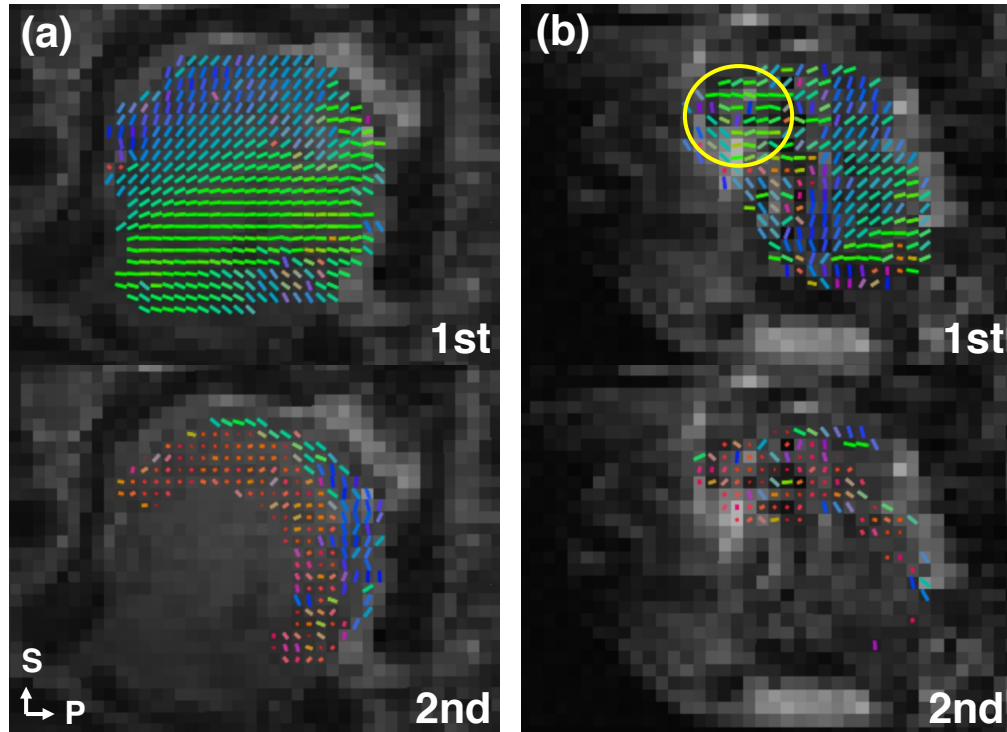
Figure 5-4. Mid-sagittal slice of the reconstructed fiber directions on data from a healthy tongue (Control #8, (a)) and a glossectomy patient (Patient #2, (b)). The results are restricted by manually delineated tongue masks. In each subject, the top row shows the first fiber direction image and the bottom row shows the second fiber direction image. The yellow circle outlines the approximate position of the flap in the patient's tongue. The fiber directions are conventionally color-coded (red: right-left, green: anterior-posterior, blue: inferior-superior). S = superior; P = posterior.

muscle fibers show similar fiber orientations, but are inconsistent, presumably due to the presence of the flap, in the right-anterior part of the tongue (yellow contour), and muscle position changes due to surgery. The whole flap region is shown in Figure 5-3(c).

5.3.2 Strain in the Line of Action in the Whole Tongue

To observe the patterns of the strain in the line of action (SLA) for each muscle, Figure 5-5 shows the mid-sagittal slice of the tongue in four subjects, including two healthy controls ((a): C1 and (b): C8) and the two patients ((c): P1 and (d): P2). The underlying

106
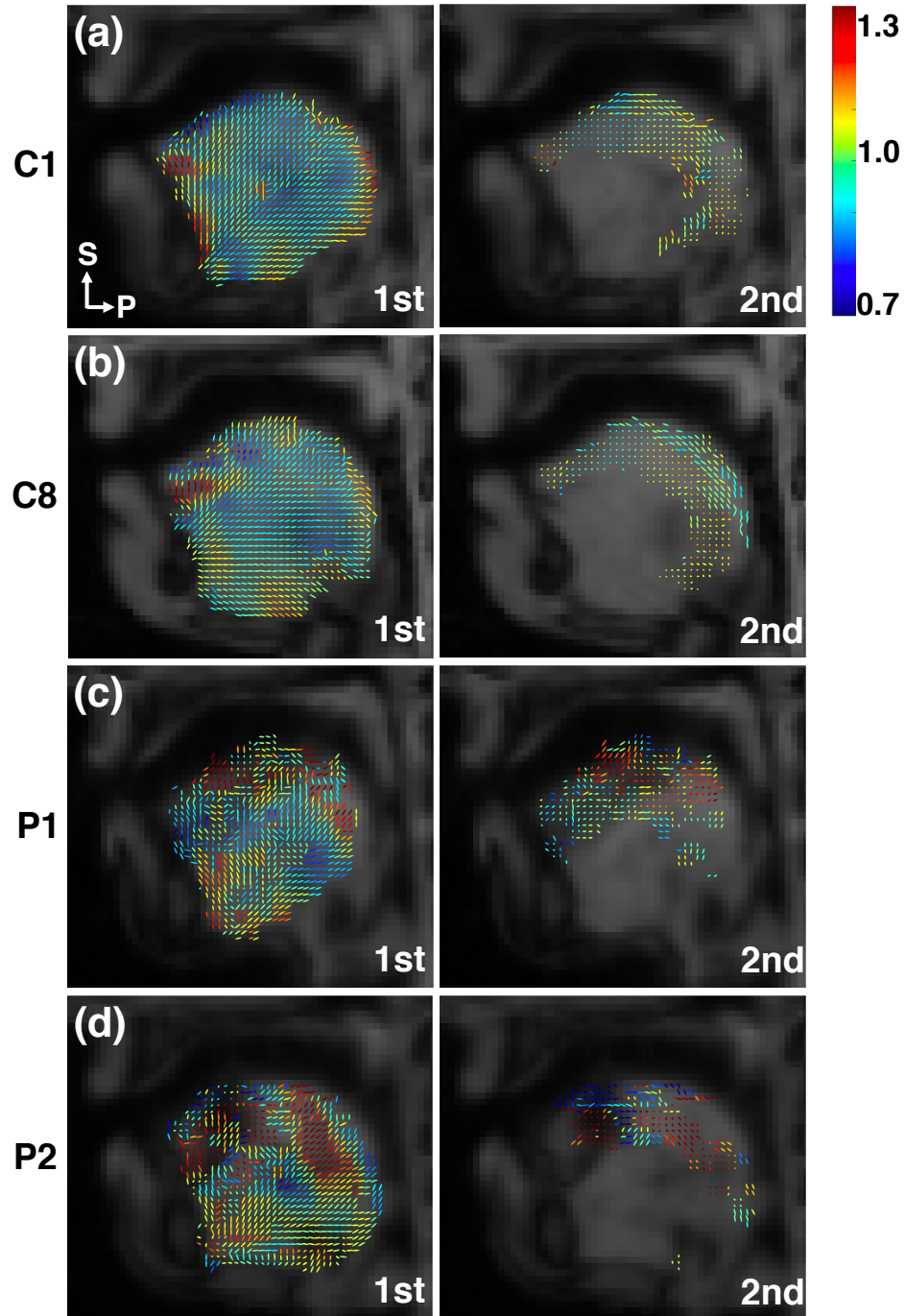
Figure 5-5. Visualization of SLAs from /ə/ to /θ/ along first (left) and second (right) tongue muscle fiber orientations. In each subfigure, the underlying image is a sagittal slice of the subject's cine image at time index $t_ə$ deformed to the cine atlas space. The colors of the fiber vectors represent the SLA values. (a) and (b) show the results on two healthy controls (C1 and C8). (c) and (d) show the results on two patients with glossectomy. S = superior; P = posterior.

image for each subject is the cine image at time index $t_\text{ə}$ deformed to the cine atlas space for easier comparison. The lines within the tongue indicate the direction of fibers that are shortening in the first direction (top) and second direction (bottom) during the motion from /ə/ to /θ/. The line colors represent the amount of compression (not fiber orientations); maximum lengthening is red, maximum shortening is blue, green is neutral, which means no compression or expansion. In each row, the left and right figures show the SLAs associated with the first and second fiber orientations, respectively.

In the SLAs along the first fiber orientation, the controls (Figures 5-5(a) and 5-5(b) top) show horizontal SLAs across the middle of the tongue and oblique ones in the upper and anterior tongue, consistent with GG muscle shortening that pulls the tongue body forward from /ə/ to /θ/. In the tip, the lower edge lengthens vertically to a large extent (red) and the upper edge shortens horizontally to a large extent (blue) consistent with activation of the superior longitudinal muscle and passive upward extension of the lower tip to reach the incisors during /θ/. The controls (Figures 5-5(a) and 5-5(b) bottom), have aqua and yellow SLA lines along the second fiber orientation that follow the surface of the tongue, consistent with the superior longitudinal fiber using no- or slight-shortening posterior to the tip and blade. The patients (Figures 5-5(c) and 5-5(d) top) show more chaotic patterns in terms of first fiber orientation, and in small regions exhibit much unconnected lengthening (red) and compression (blue) along the first fiber orientation. This is true in SLA patterns along the second fiber orientation of the patients as well (Figures 5-5(c) and 5-5(d) bottom). The SLAs in the patients reflect multiple, possibly compensatory muscle behaviors, and also deformation of the flap tissue.

To quantitatively compare the SLAs in the controls and patients, we calculated the bottom and top 10% SLA values along the first and second fiber orientations (d_1 and d_2) from /ə/ to each of the other three phonemes (/θ/, /i/, and /ŋ/) for each subject. Figure 5-6 and Figure 5-7 show the histograms of the bottom and top 10% of SLAs from /ə/ to the three phonemes for the 8 controls grouped together (green), and the two patients
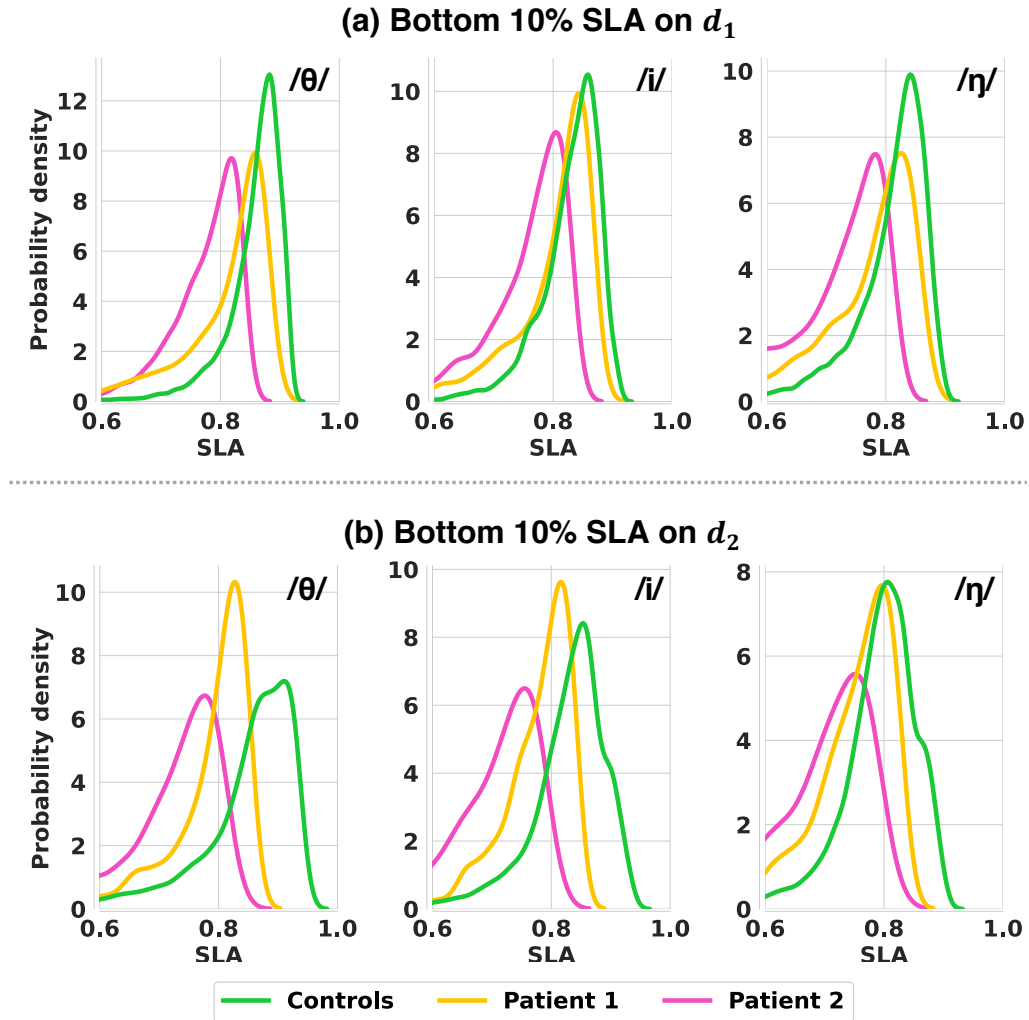
Figure 5-6. Histograms of the bottom 10% SLA values along the first and second fiber orientations from /ə/ to /θ/, /i/, and /ŋ/. The control data are grouped together.

each displayed alone (P1 in yellow, P2 in pink). The number of SLA values used to generate the histograms is in the range of 1200–2000 for each control and 700–800 for each patient along the first fiber orientations. This number is between 650–1200 for each control and 450–500 for each patient along the second fiber orientations. The data include the flap tissue as well as the muscles. The ratio of the flap volume to the whole tongue volume is 0.17 and 0.26 for Patient 1 and 2, respectively. Figures 5-6(a) and 5-6(b) show that the bottom 10% SLA along the first and second fiber orientations in the two patients are lower than those in the controls. Figures 5-7(a) and 5-7(b) show that
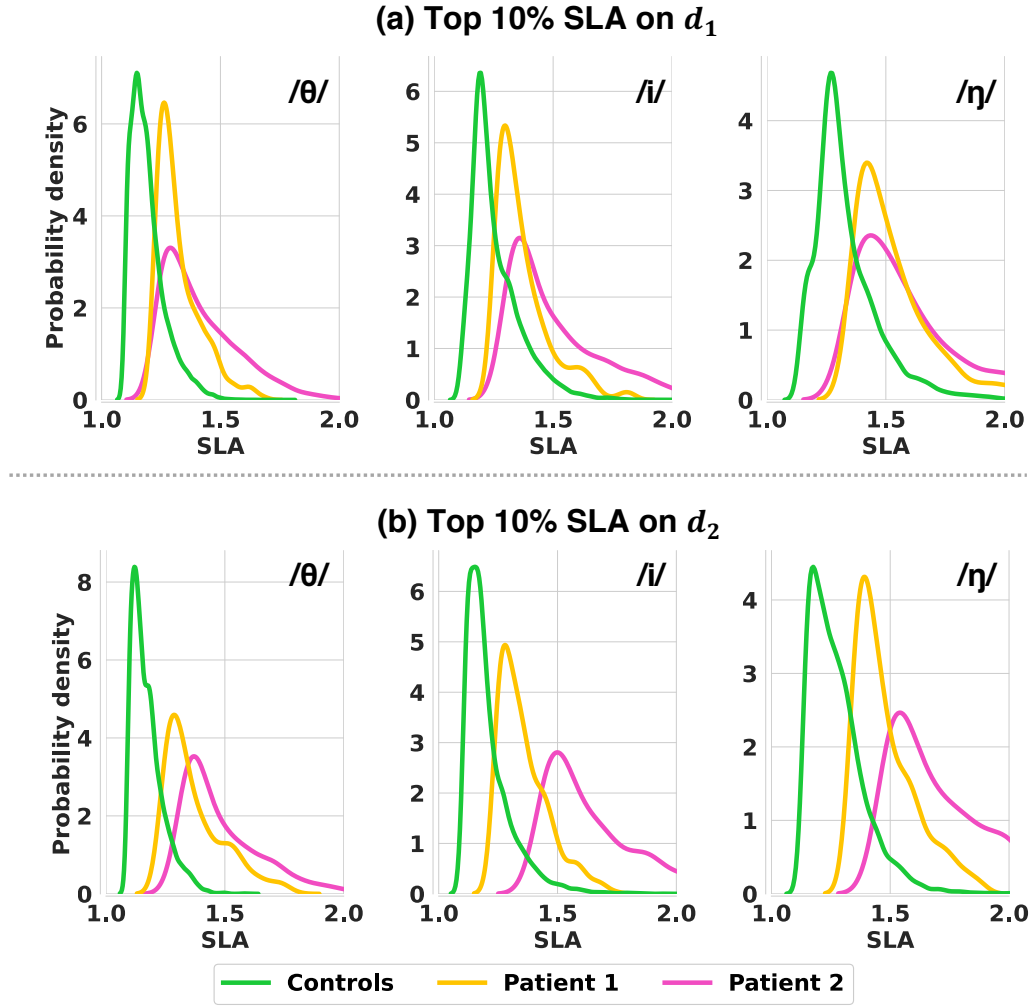
Figure 5-7. Histograms of the top 10% SLA values along the first and second fiber orientations from /ə/ to /θ/, /i/, and /ŋ/. The control data are grouped together.

the top 10% SLA in the two patients are higher than those in the controls. That is, the patients have more extreme SLA values than those of the controls.

5.3.3 Strain in the Line of Action in Muscle Groups

As described in Section 5.2.5, we used a tongue muscle mask in a T2-w MRI atlas to generate muscle masks for our subjects. The generated muscle masks were used to compute the average SLAs within the anterior and posterior GG and T muscle groups

from /ə/ to each aligned time frame. Figure 5-8 and Figure 5-9 present the time series of the average SLAs in the 10 subjects, including 8 healthy controls (C1–C8) and 2 glossectomy patients (P1 and P2). All the SLA values were calculated in the Lagrangian framework with the reference time frame being $t_ə$. SLA values greater than 1 represent extension and values less than 1 represent contraction. The lines represent the average SLAs for each subject and muscle and the similarly colored shaded regions represent the standard deviations.

In Figure 5-8, the SLAs show that the GGa muscle (blue) shortens into the /θ/ and then lengthens throughout the rest of the word for most of the controls and P1. The GGa muscle in C2 and C3 shows an initial small extension, then shortens and lengthens, but maximum shortening still occurs around the time of the /θ/. For P2, GGa maintains its resting length for part of the motion into /θ/, then lengthens.

For the Ta muscle (orange), controls C5–C8 show an oppositional pattern to GGa in which Ta shortens at or after the /θ/, consistent with inactivity during the /θ/ followed by narrowing the tongue to facilitate its upward expansion during the /i/ and /ŋ/. C2–C4 also have a Ta length greater than 1 until /i/ or /ŋ/. C1 is the outlier whose pattern strongly resembles that of P1, in which both muscles shorten into the /θ/ and then lengthen, suggesting that the Ta is used to help GGa extend the tongue tip. For P2, the average SLA values for Ta are always at or above 1, though it does shorten into the /i/ before lengthening further, like the other subjects.

In Figure 5-9, the GGp muscle (coral) shows peak shortening at the /i/ or /ŋ/ for all 10 subjects; even C2, who lengthens the muscle for most of the word, shortens it a bit before /i/, which is consistent with creating an anterior tongue root to elevate the tongue body. Shortening occurs also for /θ/ in C1, C3, C4, C6, C7, C8, and P2, consistent with pulling the tongue root anteriorly to help position the tongue tip at the teeth. The Tp muscle (green) shows increased amounts of extension toward the end of the word in C1, C2, C3, C4, C5, and C8. C6 and C7 show shortening into the /i/ and /ŋ/ (just like Ta)

111
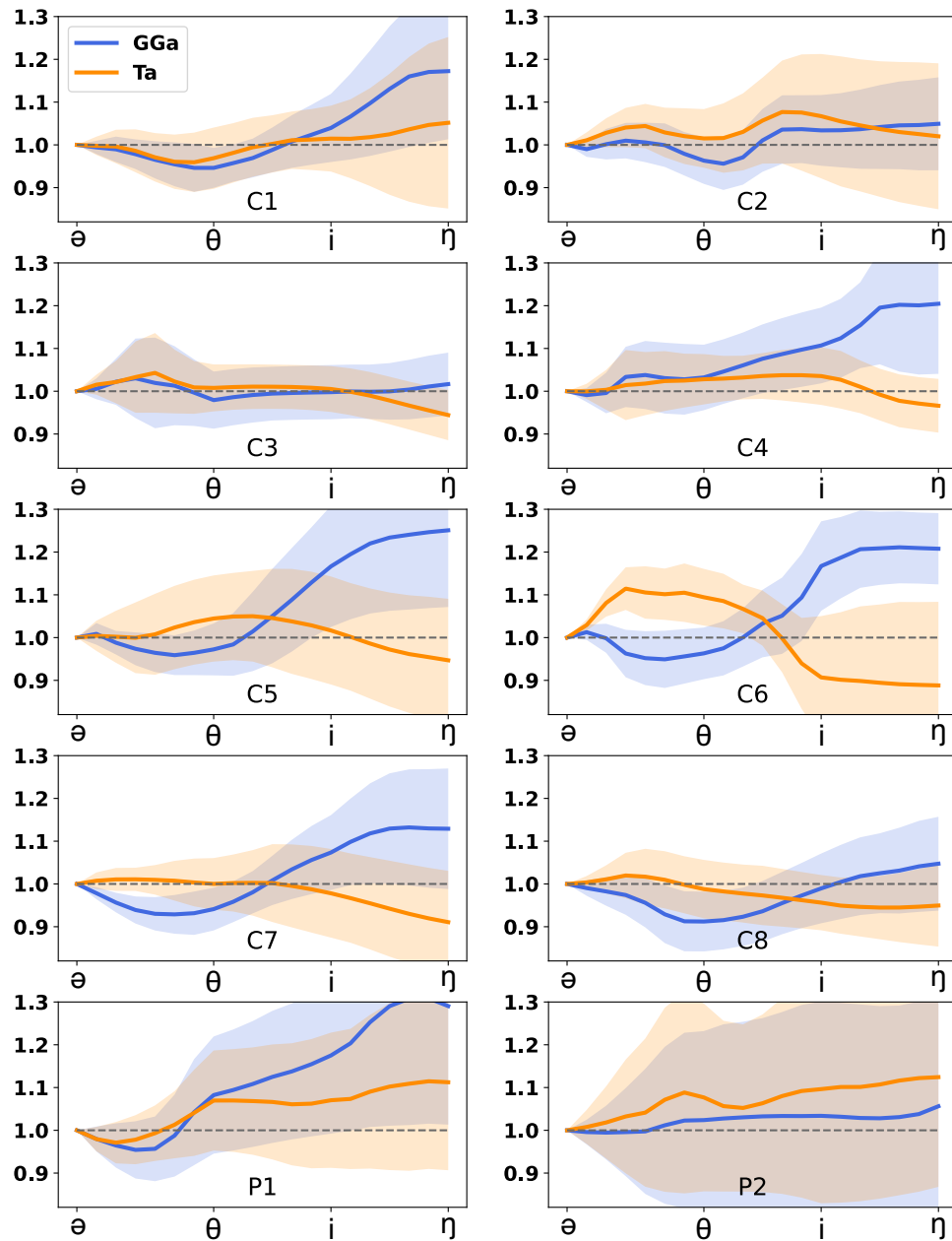
Figure 5-8. Average SLAs along the GG and T muscles calculated in the anterior part of the tongue from 8 healthy controls (C1 to C8) and 2 glossectomy patients (P1 and P2) in 20 aligned time indices pronouncing "a thing". All the SLAs were calculated in Lagrangian framework and the undeformed time frame is at /ə/. The solid lines are the mean SLA values and the shades show the standard deviations. GGa = genioglossus anterior; Ta = transverse anterior.

suggesting the muscle is behaving similarly throughout its length. In the patients, the GGp and Tp muscles show average SLA values that move similarly to each other.

5.4   Discussion

In this study, we visualized and analyzed the cooperation between certain tongue muscle groups during speech production. We performed the SLA analysis on 8 healthy controls and 2 glossectomy patients. We compared the patterns of both muscle fiber orientations and SLA patterns between controls and patients. Results showed some common patterns in the controls and differences in the patients.

The fiber orientation reconstruction network can reconstruct most fiber orientations in the tongue and predict orthogonal fiber directions. The fiber orientation matching algorithm provides clean fiber orientations in most regions of the tongue. The controls have smooth and consistent muscle fiber orientations. We can see clear fan-shaped GG muscle fiber directions in the first fiber orientation image and arc-shaped T muscle fiber orientations in the second fiber orientation image in the sagittal view, which is consistent with tongue anatomy [170]. In the patient data, we can still observe the general patterns of the GG and T muscles. The muscle fiber orientations in regions far from the flap look similar to those in the controls, while in regions near the flap, the fiber orientations look irregular (upper-left in Figure 5-4(b)). This behavior is expected because the flap region does not contain muscles.

From the visualization of the SLAs overlaid on subjects' cine images in the atlas space, we can observe and compare the way tongue muscles contract or extend at key time frames during speech production across different subjects. This visualization benefits from the fiber orientation matching algorithm as we can observe and compute the average SLAs within specific muscle groups. The two controls (Figures 5-5(a) and 5-5(b)) are shown to have similar SLA patterns from /ə/ to /θ/ with only small
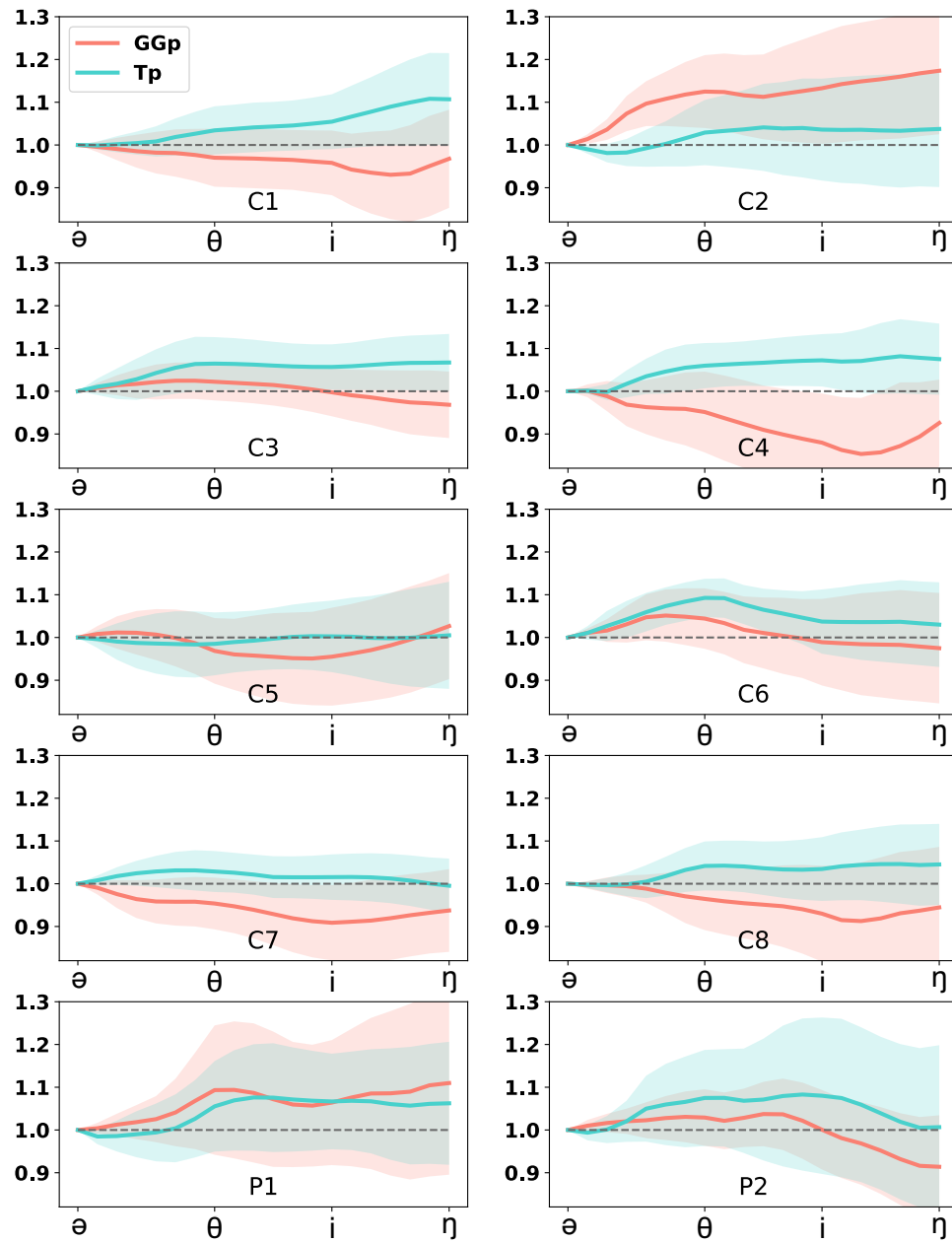
Figure 5-9. Average SLAs along the GG and T muscles calculated in the posterior part of the tongue from 8 healthy controls (C1 to C8) and 2 glossectomy patients (P1 and P2) in 20 aligned time indices pronouncing "a thing". All the SLAs were calculated in Lagrangian framework and the undeformed time frame is at /ə/. The solid lines are the mean SLA values and the shades show the standard deviations. GGp = genioglossus posterior; Tp = transverse posterior.

inconsistencies; the phoneme /θ/ is characterized by a large tongue protrusion. In order to produce /θ/, both subjects show contraction in the upper surface of the tongue tip and extension below the tip along the first fiber orientation (Figures 5-5(a) and 5-5(b) left). In the tongue body, the tongue experiences shortening along the GG muscle, which likely activates to facilitate the tongue-tooth contact. At the same time, a small extension along the T muscle is observed throughout the tongue, which shows that for /θ/, the tongue widening also occurs as the lateral tongue contacts the molars (Figures 5-5(a) and 5-5(b) right). The SLAs in the patients (Figures 5-5(c) and 5-5(d)) show different patterns. Patient 2 shows similar tip behavior as the controls: shortening above and lengthening below. However, both the patients show larger extension (dark red) than the controls in some healthy muscle tissue regions, which can be seen in the superior region of the tongue.

Quantitative measurements provide insight on the different patterns of the SLAs between controls and patients. Histograms in Figure 5-6 and Figure 5-7 show the distributions of the bottom and top 10% of SLA values for our subjects. One immediate observation is that the SLA values in the patients have wider ranges in these plots, the top 10% are higher and the bottom 10% are lower, indicating that there are more extreme SLA values in the patients' data. Patient 2 has a larger relative flap volume than patient 1 and shows a wider range of SLA values. One possible reason is that under the condition of muscle degeneration, the remaining muscles of the patient's tongue must "work" harder to compensate for the muscle loss. Compared to the controls, the patients' tongues make larger deformations during speech production showing greater compression and greater extension.

The time series plots enable us to analyze the cooperation between the GG and T muscle groups during speech production. It can be seen from Figure 5-8 and Figure 5-9 that the behaviors of the two muscles across controls show some similarity and some differences. From Figure 5-8, the time series of C4–C8 visualize some negative

correlation between the GGa and Ta muscles, as we can observe that one muscle tends to shorten while the other one extends. The remaining controls, C1, C2, and C3, show a positive correlation between the GGa and Ta muscles, as they shorten and extend fairly simultaneously in our speech task. Clearly this represents two strategies by GGa and Ta to extend the tip. Our results also show that the SLAs along the GGa muscle are lowest near t_θ in our controls; even C4 pauses GGa extension at that time. This supports the idea that in healthy controls the GGa muscle actively shortens for the /θ/ sound, possibly to ensure a local bend at the tongue blade as the superior longitudinal muscle pulls the tongue tip back. Without GGa assistance, superior longitudinal shortening would merely pull the tongue tip straight back [122].

In Figure 5-9, all the subjects except C2 and P1 shorten GGp during the motion into the /i/ and /ŋ/ gestures; the GGp muscle shows the lowest average SLA values near t_i, and t_η. This pattern documents our understanding that the generation of /i/ and /ŋ/, which require elevating the tongue body, is often accomplished by shortening the GGp muscle. In most subjects we see a negative correlation between GGp and Tp, where the GGp shortens and the Tp lengthens. This is consistent with allowing the tongue to widen as it elevates to create lingual contact with the lateral teeth (inner edge of molars) for both /i/ and /ŋ/. Subject C2, who does not shorten GGp, is the only control who does not appear to use the GGp muscle during this word, though there is a brief shortening at the /θ/ sound.

Examining the patients, their SLA patterns show different characteristics from the controls. Figure 5-8 shows both anterior muscles either shorten or stay level at the /θ/ for both patients; they appear to engage them, like the controls, to control the tip. After that neither patient shortens either anterior muscle. Figure 5-9 shows positively correlated SLA patterns for the posterior muscles, which could indicate reduction in motor control, or the flaps might inhibit tongue deformation. Both patients extend both muscles through the /θ/, showing no strong usage of the posterior muscles for /θ/, with stable length (P1),

or shortening (P2) into the /i/ and /ŋ/.

In P2, GGa, Ta, and Tp show extension in general. However, the GGa shortens before the /i/, and the GGp shortens toward the end of the phrase, reflecting the usage of these muscles by P2 but less so by P1. The standard deviation of the SLAs in the two patients are larger than those of the controls. This is unsurprising as patients are often more variable than controls. The two muscle groups also tend to extend and compress more locally compared to the controls, as seen in Figure 5-5. It should be remembered that the patients are at all times moving an incompressible, inert flap, which affects the shapes, the durations, and the variability of their gestures.

With our strain analysis pipeline, we can observe the patterns of cooperation between muscle groups evidenced by simultaneous and antagonistic behaviors within each subject during speech production. We can also see that some patterns are consistent in control subjects, although there exist unique patterns in individuals. We also see that glossectomy patients show different SLA patterns than controls. The comparison between the controls and patients could provide some insights about the adaptive behaviors in speech production of the glossectomy patients.

One limitation of the study is that there is very limited previous work on this topic. Therefore, the observations and conclusions in this work are hard to verify with previous research. The current findings from the results are only based on a small cohort and a single speech task. More concrete conclusions can be made as further research develops.

5.5   Summary

In this study, we presented a workflow to analyze strain in the line of action associated with tongue muscle fiber orientations in speech sound production using multiple MRI modalities. We used the fiber orientation reconstruction network and the fiber orienta-

tion matching algorithm described in Chapter 4 to reconstruct consistent muscle fiber orientations from diffusion MRI. Tagged MRI is used to quantify the tongue motion in speech production. Strains in the line of action are computed within the tongue muscle masks. We performed analysis on a cohort of 8 healthy controls and 2 glossectomy patients. Visual evidence of correlation between two muscle groups is presented in the analysis. Results show consistency of muscle behaviors among some healthy controls during speech production. Patients tend to have somewhat different strain patterns than the controls. The proposed pipeline provides a solution to quantitatively analyze the cooperation between tongue muscles during speech production.

Chapter 6

Conclusions and Future Work

6.1   Summary

In this dissertation, we developed a set of deep learning methods for MRI analysis using different image modalities. In Chapter 2, we presented a modified 3D U-Net for ventricle parcellation of brain with ventriculomegaly from T1-w MRI. In Chapter 3, we proposed a multi-channel 3D U-Net to parcellate the thalamus from T1-w MRI and diffusion MRI, where we explored the impact of an MR image harmonization on the segmentation network. In Chapter 4, we developed a 3D patch-based CNN to directly estimate the crossing muscle fiber orientations in the human tongue using diffusion MRI. We also proposed a fiber orientation matching algorithm to refine the assignments of the fiber reconstruction. In Chapter 5, we presented a pipeline to analyze the strain in the line of action along the muscle fiber orientations during speech production. In the following, we summarize each topic and discuss potential future work.

6.2 Brain Ventricle Parcellation using T1-w MRI

6.2.1 Key Ideas and Results

1. We developed a fully automatic deep learning method, VParNet, to parcellate the brain ventricle into its sub-compartments from T1-w MRI. VParNet is modified from a 3D U-Net and incorporates a residual block in the encoder. It can be applied to both healthy subjects and patients with ventriculomegaly.

2. VParNet was evaluated on two data sets with manual delineations. It achieved state-of-the-art performance on a healthy data set and significantly better performance than the comparing methods on a normal pressure hydrocephalus data set. It achieved a mean Dice similarity coefficient of 0.895 for the whole brain ventricle on the healthy data set and 0.973 on the hydrocephalus data.

3. We carried out a small study to apply VParNet on subjects treated by shunt surgery. We did not observe major failures in the parcellation results.

6.2.2 Future Work

In medical image segmentation, labeled data is always limited. Besides the standard data augmentation methods, an anatomical prior could be incorporated into the network for better segmentation.

Although the training data of VParNet include MRIs from patients with enlarged ventricles, post-shunt-surgery data with implanted valves show different properties from the training data. Therefore, it is expected to observe degraded performance of ventricle parcellation when applying VParNet to these data. We can observe that VParNet mislabeled some dark regions near the implanted valve as ventricle labels. To improve the accuracy of the parcellation in these data, some post-processing steps can be used

to remove small false positive labels. Another possible improvement is to incorporate a topological constraint in the proposed method, since each sub-compartment of the ventricular system is assumed to be a single connected component. We can also try some more advantage deep learning techniques. For example, attention modules can be incorporated, which might improve the segmentation performance [171, 172].

6.3   Brain thalamus parcellation using multimodal MRI

6.3.1   Key Ideas and Results

1. We developed a multi-channel deep learning method to segment the brain thalamus and parcellate each hemisphere into six thalamic nuclei from multimodal MRIs. The proposed network is modified from a 3D U-Net and the input to the network includes a T1-w MR image and features derived from DTI.

2. We incorporated an MR image harmonization algorithm as a pre-processing step for the T1-w MR image. The training and testing data were harmonized to the same target data set. Therefore, the contrast variation between the training and testing data was reduced to improve the generalization ability of the parcellation network.

3. We did a cross validation on the proposed network on a set of 21 subjects with manual delineations. It produced a mean Dice similarity coefficient of 0.861 and 95% Hausdorff distance of 1.440 mm for the whole thalamus segmentation. The proposed method was compared against widely used segmentation algorithms and achieved significantly better performance on thalamus segmentation.

4. When evaluating on the same data set, the proposed network trained on harmonized data achieved competitive results on thalamus segmentation and parcellation

compared to the same network trained on unharmonized data. While evaluating on an unseen data set, image harmonization improved the segmentation accuracy of the proposed network. The experiments showed that MR image harmonization can match the contrast between the training and testing data while maintaining the underlying anatomy in the MR image. The network trained on harmonized data provides the potential to process large amounts of data from other data sets, even in the absence of new training data.

6.3.2 Future Work

Deep learning based segmentation with limited training data is challenging in medical image analysis. Large number of training data with different contrasts will help the network generalize better. However, it is very time-consuming and requires anatomy knowledge to delineate brain structures. Therefore, we incorporated an MR image harmonization algorithm as pre-processing step to reduce contrast variation. As there exist more data without manual delineations, semi-supervised algorithms could be explored to utilize unlabeled data to improve the network performance.

New features could be explored and incorporated in the thalamus parcellation network. For example, since each thalamic nucleus has connections to different parts of the brain, connectivity between the thalamus and the brain can be used to improve parcellation of the thalamus. One possible way to measure the connectivity is to perform probabilistic tractography in the brain [173]. White-matter-nulled MPRAGE has been investigated to visualize the anatomy of the thalamus [174, 175]. If these data were acquired, additional features that provide better contrast of thalamic nuclei could be added to the proposed network.

6.4 Reconstruction of Tongue Muscle Fiber Orientations using Diffusion MRI

6.4.1 Key Ideas and Results

1. We developed a patch-based CNN algorithm to reconstruct the tongue muscle fiber orientations from diffusion MRI. A separation loss was proposed to encourage the network to predict orthogonal crossing fiber orientations, which is an important property of the tongue muscles.

2. We presented a fiber orientation matching algorithm to refine the assignment of the reconstructed fiber orientations. The algorithm formed a quadratic unconstrained binary optimization framework to find optimal assignment of the fiber orientations for each voxel in the tongue region.

3. The proposed fiber orientation reconstruction network was evaluated on synthetic and post-mortem tongue data. The network achieved a success rate of 0.717 and mean angular error of 5.9° in single fiber region and 13.0° in crossing fiber region in a synthetic tongue data with SNR = 20. The proposed method shows better performance than the comparing methods in regions with crossing fiber orientations. Visual comparison on the real tongue data shows that the proposed method produced cleaner fiber orientations.

4. The proposed fiber matching algorithm was quantitatively evaluated on synthetic tongue muscle fibers. The method switched the fiber assignments in regions with swapped fiber orientations and provided more consistent fiber orientations.

6.4.2 Future Work

Accurate reconstruction of tongue muscle fiber orientations is challenging. Currently, the orthogonal characteristic of the crossing fiber orientations is encouraged by the separation loss. Further improvement could incorporate some distribution prior to the network.

One limitation of the proposed fiber orientation reconstruction network is that it assumes there is at least one fiber orientation in each voxel, which may not be true in some regions. This can be addressed by adding the fractional anisotropy (FA) map as a constraint. Since low FA values indicate isotropic water motion in the underlying tissue, no fiber orientation should be assigned. More sophisticated network structures can also be explored to improve the network performance.

The proposed fiber matching algorithm was specifically designed for tongue fiber orientations as it assumes the the crossing fiber orientations are nearly orthogonal. Therefore, it cannot be generalized to white matter tracts in the brain. Investigating new objective function to make it work in brain tissue can also be an interesting topic.

6.5 Tongue Strain Analysis Pipeline

6.5.1 Key Ideas and Results

1. We developed a pipeline to analyze the strains in the line of action (SLA) along the tongue muscle fiber orientations during speech generation. Multimodal MRIs are used in the processing. Diffusion MRI is used to reconstruct the tongue muscle fiber orientations. Tagged and cine MRIs are used to obtain the tongue motion during speech.

2. All the subjects' tongue motion fields were temporally and spatially aligned to an

atlas space. The reconstructed fiber orientations were transformed to the motion space to compute strain along the local fibers.

3. We performed an analysis on a cohort of 8 healthy controls and 2 post-partial-glossectomy patients and achieved both qualitative and quantitative results. Results showed some common patterns in the controls and differences in the patients. The proposed pipeline has the potential to aid researchers in better understanding the cooperation between muscle groups in speech production and in learning about the adaptive behaviors of patients with glossectomies.

6.5.2 Future Work

Currently, there exists very limited previous work on this topic. More subjects and more speech tasks can be explored to draw concrete conclusions on the cooperation patterns between muscle groups.

The processing pipeline can be further developed. For example, the alignment between the diffusion MRI in the static state and the cine MRI in the motion state is achieved by registration between tongue masks. In the future, the intensity images can be incorporated to improve the registration accuracy inside the tongue region. The pipeline utilizes multiple MRI modalities for analysis, and to acquire these images requires a long scanning time for the subject. Tagged-to-cine MRI synthesis algorithm [176] has been developed to generate cine MRI based on tagged MRI. The pipeline can take advantage of this algorithm to acquire anatomical information from the tagged data, and therefore avoid using the cine data and reduce the scanning time.

6.6   Conclusion

In this dissertation, we have developed deep learning algorithms to analyze multimodal MRIs. We developed algorithms for brain structures segmentation and tongue muscle fiber reconstruction. Each algorithm has been validated using a set of experiments. In practice, the brain segmentation algorithms have been applied in large amounts of data for statistical studies. This research also includes a wide range of other topics in the field of medical image processing, including image registration, image harmonization, atlas construction, etc. We hope that the work presented in this dissertation can provide more insights in the field and have a real impact in medical science.

References

1. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* **56,** 907–922 (2011).

2. Fischl, B. FreeSurfer. *NeuroImage* **62,** 774–781 (2012).

3. Stone, M., Langguth, J. M., Woo, J., Chen, H. & Prince, J. L. Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis. *Journal of Speech, Language, and Hearing Research* **57.** Publisher: ASHA, 707–717 (2014).

4. Xing, F., Woo, J., Lee, J., Murano, E. Z., Stone, M. & Prince, J. L. Analysis of 3-D tongue motion from tagged and cine magnetic resonance images. *Journal of Speech, Language, and Hearing Research* **59.** Publisher: ASHA, 468–479 (2016).

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

6. Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering* **19,** 221 (2017).

7. Nair, V. & Hinton, G. E. *Rectified linear units improve restricted boltzmann machines* in *Proceedings of the 27th International Conference on Machine Learning* (2010), 807–814.

8. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* **15,** 1929–1958 (2014).

9. Ioffe, S. & Szegedy, C. *Batch normalization: Accelerating deep network training by reducing internal covariate shift* in *International Conference on Machine Learning* (2015), 448–456.

10. Ulyanov, D., Vedaldi, A. & Lempitsky, V. S. *Improved Texture Networks: Maximizing Quality and Diversity in Feed-forward Stylization and Texture Synthesis* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), 6924–6932.

11. Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural networks* **2,** 359–366 (1989).

12. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323,** 533–536 (1986).

13. Amari, S.-i. Backpropagation and stochastic gradient descent method. *NeuroComputing* **5,** 185–196 (1993).

14. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

15. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative adversarial nets. *Advances in Neural Information Processing Systems* **27,** 2672–2680 (2014).

17. Taylor, L. & Nitschke, G. *Improving deep learning with generic data augmentation* in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (2018), 1542–1547.

18. Ronneberger, O., Fischer, P. & Brox, T. *U-net: Convolutional networks for biomedical image segmentation* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015), 234–241.

19. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. *3D U-Net: learning dense volumetric segmentation from sparse annotation* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016), 424–432.

20. Shao, M., Han, S., Carass, A., Li, X., Blitz, A. M., Prince, J. L. & Ellingsen, L. M. *Shortcomings of Ventricle Segmentation Using Deep Convolutional Networks* in *Understanding and Interpreting Machine Learning in Medical Image Computing Applications* (2018), 79–86.

21. Shao, M., Han, S., Carass, A., Li, X., Blitz, A. M., Shin, J., Prince, J. L. & Ellingsen, L. M. Brain ventricle parcellation using a deep neural network: Application to patients with ventriculomegaly. *NeuroImage: Clinical* **23,** 101871 (2019).

22. Yamada, S., Ishikawa, M. & Yamamoto, K. Optimal diagnostic indices for idiopathic normal pressure hydrocephalus based on the 3D quantitative volumetric analysis for the cerebral ventricle and subarachnoid space. *American Journal of Neuroradiology* **36,** 2262–2269 (2015).

23. Benedetto, N., Gambacciani, C., Aquila, F., Di Carlo, D. T., Morganti, R. & Perrini, P. A new quantitative method to assess disproportionately enlarged subarachnoid space (DESH) in patients with possible idiopathic normal pressure hydrocephalus: The SILVER index. *Clinical Neurology and Neurosurgery* **158,** 27–32 (2017).

24. Shao, M., Zuo, L., Carass, A., Zhuo, J., Gullapalli, R. P. & Prince, J. L. *Evaluating the impact of MR image harmonization on thalamus deep network segmentation* in *Medical Imaging 2022: Image Processing* **12032** (2022), 115–121.

25. Ryan, N. S., Keihaninejad, S., Shakespeare, T. J., Lehmann, M., Crutch, S. J., Malone, I. B., Thornton, J. S., Mancini, L., Hyare, H., Yousry, T., *et al.* Magnetic resonance imaging evidence for presymptomatic change in thalamus and caudate in familial Alzheimer's disease. *Brain* **136,** 1399–1414 (2013).

26. Cifelli, A., Arridge, M., Jezzard, P., Esiri, M. M., Palace, J. & Matthews, P. M. Thalamic neurodegeneration in multiple sclerosis. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **52,** 650–653 (2002).

27. Zhuo, J., Jiang, L., Sours Rhodes, C., Roys, S., Shanmuganathan, K., Chen, H., Prince, J. L., Badjatia, N. & Gullapalli, R. P. Early Stage Longitudinal Subcortical

Volumetric Changes following Mild Traumatic Brain Injury. *Brain injury* **35,** 725–733 (2021).

28. Zuo, L., Dewey, B. E., Liu, Y., He, Y., Newsome, S. D., Mowry, E. M., Resnick, S. M., Prince, J. L. & Carass, A. Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. *NeuroImage* **243,** 118569 (2021).

29. Shao, M., Carass, A., Gomez, A. D., Zhuo, J., Liang, X., Stone, M. & Prince, J. L. *Direct reconstruction of crossing muscle fibers in the human tongue using a deep neural network* in *Computational Diffusion MRI* (2021), 69–80.

30. Voskuilen, L., Mazzoli, V., Oudeman, J., Balm, A. J., van der Heijden, F., Froeling, M., de Win, M. M., Strijkers, G. J., Smeele, L. E. & Nederveen, A. J. Crossing muscle fibers of the human tongue resolved in vivo using constrained spherical deconvolution. *Journal of Magnetic Resonance Imaging* **50,** 96–105 (2019).

31. Shao, M., Gomez, A. D., Zhuo, J., Liang, X., Stone, M., Carass, A. & Prince, J. L. *Reconstruction and refinement of crossing muscle fibers in the human tongue* in *Medical Imaging 2021: Image Processing* **11596** (2021), 424–430.

32. Hiiemae, K. M. & Palmer, J. B. Tongue movements in feeding and speech. *Critical Reviews in Oral Biology & Medicine* **14,** 413–429 (2003).

33. Woo, J., Xing, F., Prince, J. L., Stone, M., Green, J. R., Goldsmith, T., Reese, T. G., Wedeen, V. J. & El Fakhri, G. Differentiating post-cancer from healthy tongue muscle coordination patterns during speech using deep learning. *The Journal of the Acoustical Society of America* **145,** EL423–EL429 (2019).

34. Xing, F., Woo, J., Gomez, A. D., Pham, D. L., Bayly, P. V., Stone, M. & Prince, J. L. Phase vector incompressible registration algorithm for motion estimation from tagged magnetic resonance images. *IEEE transactions on medical imaging* **36,** 2116–2128 (2017).

35. Nolte, J. *The human brain: an introduction to its functional anatomy* (Elsevier, 2009).

36. Bradley, W. CSF Flow in the Brain in the Context of Normal Pressure Hydrocephalus. *American Journal of Neuroradiology* **36,** 831–838 (2015).

37. McGirt, M. J., Woodworth, G., Coon, A. L., Thomas, G., Williams, M. A. & Rigamonti, D. Diagnosis, treatment, and analysis of long-term outcomes in idiopathic normal-pressure hydrocephalus. *Neurosurgery* **57,** 699–705 (2005).

38. Olivero, M. *Is It Really Dementia? Maybe Not.* https://health.usnews.com/health-news/patient-advice/articles/2015/02/05/is-it-dementia-or-normal-pressure-hydrocephalus. 2015.

39. Jaraj, D., Rabiei, K., Marlow, T., Jensen, C., Skoog, I. & Wikkelsø, C. Prevalence of idiopathic normal-pressure hydrocephalus. *Neurology* **82,** 1449–1454 (2014).

40. Stein, S. C., Burnett, M. G. & Sonnad, S. S. Shunts in normal-pressure hydrocephalus: do we place too many or too few? *Journal of Neurosurgery* **105,** 815–822 (2006).

41. Panagiotopoulos, V., Konstantinou, D., Kalogeropoulos, A. & Maraziotis, T. The predictive value of external continuous lumbar drainage, with cerebrospinal fluid outflow controlled by medium pressure valve, in normal pressure hydrocephalus. *Acta Neurochirurgica* **147,** 953–958 (2005).

42. Leinonen, V., Koivisto, A. M., Alafuzoff, I., Pyykkö, O. T., Rummukainen, J., von Und Zu Fraunberg, M., Jääskeläinen, J. E., Soininen, H., Rinne, J. & Savolainen, S. Cortical brain biopsy in long-term prognostication of 468 patients with possible normal pressure hydrocephalus. *Neurodegenerative Diseases* **10,** 166–169 (2012).

43. Mori, E., Ishikawa, M., Kato, T., Kazui, H., Miyake, H., Miyajima, M., Nakajima, M., Hashimoto, M., Kuriyama, N. & Tokuda, T. Guidelines for management of idiopathic normal pressure hydrocephalus: second edition. *Neurologia Medico-chirurgica* **52,** 775–809 (2012).

44. Kang, K., Ko, P.-W., Jin, M., Suk, K. & Lee, H.-W. Idiopathic normal-pressure hydrocephalus, cerebrospinal fluid biomarkers, and the cerebrospinal fluid tap test. *Journal of Clinical Neuroscience* **21,** 1398–1403 (2014).

45. Wikkelsø, C., Hellström, P., Klinge, P. M., Tans, J. T. J. & iNPH Multicentre Study Group, E. The European iNPH Multicentre Study on the predictive values of resistance to CSF outflow and the CSF Tap Test in patients with idiopathic normal pressure hydrocephalus. *Journal of Neurology, Neurosurgery, and Psychiatry* **84,** 562–568 (2013).

46. Shiee, N., Bazin, P.-L., Cuzzocreo, J. L., Blitz, A. & Pham, D. L. Segmentation of brain images using adaptive atlases with application to ventriculomegaly. *Information Processing in Medical Imaging* **22,** 1–12 (2011).

47. Roy, S., He, Q., Sweeney, E., Carass, A., Reich, D. S., Prince, J. L. & Pham, D. L. Subject specific sparse dictionary learning for atlas-based brain MRI segmentation. *IEEE Journal of Biomedical and Health Informatics* **19,** 1598–1609 (2015).

48. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* **9,** 179–194 (1999).

49. Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33,** 341–355 (2002).

50. Ledig, C., Heckemann, R. A., Hammers, A., Lopez, J. C., Newcombe, V. F., Makropoulos, A., Lötjönen, J., Menon, D. K. & Rueckert, D. Robust whole-brain segmentation: Application to traumatic brain injury. *Medical Image Analysis* **21,** 40–58 (2015).

51. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* **12,** 26–41 (2008).

52. Wang, H. & Yushkevich, P. Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation. *Frontiers in Neuroinformatics* **7,** 27 (2013).

53. Manjón, J. V. & Coupé, P. volBrain: An online MRI brain volumetry system. *Frontiers in Neuroinformatics* **10,** 30 (2016).

54. Kempton, M. J., Underwood, T. S., Brunton, S., Stylios, F., Schmechtig, A., Ettinger, U., Smith, M. S., Lovestone, S., Crum, W. R. & Frangou, S. A comprehensive testing protocol for MRI neuroanatomical segmentation techniques: Evaluation of a novel lateral ventricle segmentation method. *NeuroImage* **58,** 1051–1059 (2011).

55. Bazin, P.-L. & Pham, D. L. Homeomorphic brain image segmentation with topological and statistical atlases. *Medical Image Analysis* **12,** 616–625 (2008).

56. Cardoso, M. J., Clarkson, M. J., Ridgway, G. R., Modat, M., Fox, N. C., Ourselin, S., Initiative, A. D. N., *et al.* LoAd: a locally adaptive cortical segmentation algorithm. *NeuroImage* **56,** 1386–1397 (2011).

57. Doshi, J., Erus, G., Ou, Y., Resnick, S. M., Gur, R. C., Gur, R. E., Satterthwaite, T. D., Furth, S., Davatzikos, C., Initiative, A. N., *et al.* MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *NeuroImage* **127,** 186–195 (2016).

58. Shattuck, D. W. & Leahy, R. M. BrainSuite: an automated cortical surface identification tool. *Medical Image Analysis* **6,** 129–142 (2002).

59. Asman, A. J. & Landman, B. A. Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis* **17,** 194–208 (2013).

60. Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C. & Yushkevich, P. A. Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35,** 611–623 (2013).

61. Ellingsen, L. M., Roy, S., Carass, A., Blitz, A. M., Pham, D. L. & Prince, J. L. *Segmentation and labeling of the ventricular system in normal pressure hydrocephalus using patch-based tissue classification and multi-atlas labeling* in *Medical Imaging 2016: Image Processing* (2016), 116–122.

62. Carass, A., Shao, M., Li, X., Dewey, B. E., Blitz, A. M., Roy, S., Pham, D. L., Prince, J. L. & Ellingsen, L. M. *Whole brain parcellation with pathology: Validation on*

ventriculomegaly patients in *International Workshop on Patch-Based Techniques in Medical Imaging* (2017), 20–28.

63. Shao, M., Carass, A., Li, X., Dewey, B. E., Blitz, A. M., Prince, J. L. & Ellingsen, L. M. *Multi-atlas segmentation of the hydrocephalus brain using an adaptive ventricle atlas* in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging* (2018), 100–106.

64. Babalola, K. O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., Smith, S., Cootes, T., Jenkinson, M. & Rueckert, D. An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *NeuroImage* **47,** 1435–1447 (2009).

65. Brant-Zawadzki, M., Gillan, G. D. & Nitz, W. R. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. *Radiology* **182,** 769–775 (1992).

66. Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C. & Buckner, R. L. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* **19,** 1498–1507 (2007).

67. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. & Gee, J. C. N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging* **29,** 1310–1320 (2010).

68. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage,* S102 (2009).

69. Roy, S., Butman, J. A., Pham, D. L., Initiative, A. D. N., *et al.* Robust skull stripping using multiple MR image contrasts insensitive to pathology. *NeuroImage* **146,** 132–147 (2017).

70. Kayalibay, B., Jensen, G. & van der Smagt, P. CNN-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056* (2017).

71. He, K., Zhang, X., Ren, S. & Sun, J. *Identity mappings in deep residual networks* in *European Conference on Computer Vision* (2016), 630–645.

72. Maas, A. L., Hannun, A. Y. & Ng, A. Y. *Rectifier nonlinearities improve neural network acoustic models* in *Proceedings of International Conference on Machine Learning* **30** (2013), 3.

73. He, K., Zhang, X., Ren, S. & Sun, J. *Deep residual learning for image recognition* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), 770–778.

74. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. *Going deeper with convolutions* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), 1–9.

75. Dolz, J., Desrosiers, C. & Ayed, I. B. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage* **170,** 456–470 (2018).

76. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26,** 297–302 (1945).

77. Dubuisson, M.-P. & Jain, A. K. *A modified Hausdorff distance for object matching* in *Proceedings of 12th International Conference on Pattern Recognition* (1994), 566–568.

78. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin* **1,** 80–83 (1945).

79. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* **18,** 50–60 (1947).

80. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C. M., Emberton, M., *et al.* Weakly-supervised convolutional neural networks for multimodal image registration. *Medical Image Analysis* **49,** 1–13 (2018).

81. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J. & Dalca, A. V. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging* **38,** 1788–1800 (2019).

82. De Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M. & Išgum, I. A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis* **52,** 128–143 (2019).

83. Sherman, S. M. & Guillery, R. W. *Exploring the thalamus* (Elsevier, 2001).

84. Mai, J. K. & Majtanik, M. Toward a common terminology for the thalamus. *Frontiers in Neuroanatomy* **12,** 114 (2019).

85. Iglesias, J. E., Insausti, R., Lerma-Usabiaga, G., Bocchetta, M., Van Leemput, K., Greve, D. N., Van der Kouwe, A., Fischl, B., Caballero-Gaudes, C., Paz-Alonso, P. M., *et al.* A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *NeuroImage* **183,** 314–326 (2018).

86. Morel, A., Magnin, M. & Jeanmonod, D. Multiarchitectonic and stereotactic atlas of the human thalamus. *Journal of Comparative Neurology* **387,** 588–630 (1997).

87. Morel, A. *Stereotactic atlas of the human thalamus and basal ganglia* (CRC Press, 2007).

88. Child, N. D. & Benarroch, E. E. Anterior nucleus of the thalamus: functional organization and clinical implications. *Neurology* **81,** 1869–1876 (2013).

89. Ouhaz, Z., Fleming, H. & Mitchell, A. S. Cognitive functions and neurodevelopmental disorders involving the prefrontal cortex and mediodorsal thalamus. *Frontiers in Neuroscience* **12,** 33 (2018).

90. Kaas, J. H. & Lyon, D. C. Pulvinar contributions to the dorsal and ventral streams of visual processing in primates. *Brain Research Reviews* **55,** 285–296 (2007).

91. Cudeiro, J. & Sillito, A. M. Looking back: corticothalamic feedback and early visual processing. *Trends in Neurosciences* **29,** 298–306 (2006).

92. Stough, J. V., Glaister, J., Ye, C., Ying, S. H., Prince, J. L. & Carass, A. *Automatic method for thalamus parcellation using multi-modal feature classification* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2014), 169–176.

93. Glaister, J., Carass, A., Stough, J. V., Calabresi, P. A. & Prince, J. L. *Thalamus parcellation using multi-modal feature classification and thalamic nuclei priors* in *Medical Imaging 2016: Image Processing* (2016), 937–942.

94. Battistella, G., Najdenovska, E., Maeder, P., Ghazaleh, N., Daducci, A., Thiran, J.-P., Jacquemont, S., Tuleasca, C., Levivier, M., Bach Cuadra, M., *et al.* Robust thalamic nuclei segmentation method based on local diffusion magnetic resonance properties. *Brain Structure and Function* **222,** 2203–2216 (2017).

95. Su, J. H., Thomas, F. T., Kasoff, W. S., Tourdias, T., Choi, E. Y., Rutt, B. K. & Saranathan, M. Thalamus Optimized Multi Atlas Segmentation (THOMAS): fast, fully automated segmentation of thalamic nuclei from structural MRI. *NeuroImage* **194,** 272–282 (2019).

96. Ziyan, U., Tuch, D. & Westin, C.-F. *Segmentation of thalamic nuclei from DTI using spectral clustering* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2006), 807–814.

97. Mang, S. C., Busza, A., Reiterer, S., Grodd, W., Klose & Uwe. Thalamus segmentation based on the local diffusion direction: a group study. *Magnetic Resonance in Medicine* **67,** 118–126 (2012).

98. Le Bihan, D. & Breton, E. Imagerie de diffusion in vivo par résonance magnétique nucléaire. *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre* **301,** 1109–1112 (1985).

99. Merboldt, K.-D., Hanicke, W. & Frahm, J. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance (1969)* **64,** 479–486 (1985).

100. Basser, P. J., Mattiello, J. & LeBihan, D. MR diffusion tensor spectroscopy and imaging. *Biophysical journal* **66,** 259–267 (1994).

101. Le Bihan, D., Mangin, J.-F., Poupon, C., Clark, C. A., Pappata, S., Molko, N. & Chabriat, H. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **13,** 534–546 (2001).

102. Descoteaux, M. High angular resolution diffusion imaging (HARDI). *Wiley Encyclopedia of Electrical and Electronics Engineering,* 1–25 (1999).

103. Basser, P. J., Mattiello, J. & LeBihan, D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* **103,** 247–254 (1994).

104. Knutsson, H. Producing a continuous and distance preserving 5-D vector representation of 3-D orientation. *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management,* 175–182 (1985).

105. Fan, X., Thompson, M., Bogovic, J. A., Bazin, P.-L. & Prince, J. L. *A novel contrast for DTI visualization for thalamus delineation* in *Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling* (2010), 977–985.

106. Deoni, S. C., Rutt, B. K., Parrent, A. G. & Peters, T. M. Segmentation of thalamic nuclei using a modified k-means clustering algorithm and high-resolution quantitative magnetic resonance imaging at 1.5 T. *NeuroImage* **34,** 117–126 (2007).

107. Majdi, M. S., Keerthivasan, M. B., Rutt, B. K., Zahr, N. M., Rodriguez, J. J. & Saranathan, M. Automated thalamic nuclei segmentation using multi-planar cascaded convolutional neural networks. *Magnetic resonance imaging* **73,** 45–54 (2020).

108. Duan, Y., Li, X. & Xi, Y. Thalamus segmentation from diffusion tensor magnetic resonance imaging. *International Journal of Biomedical Imaging* **2007** (2007).

109. He, Y., Carass, A., Zuo, L., Dewey, B. E. & Prince, J. L. *Self domain adapted network* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2020), 437–446.

110. Dewey, B. E., Zuo, L., Carass, A., He, Y., Liu, Y., Mowry, E. M., Newsome, S., Oh, J., Calabresi, P. A. & Prince, J. L. *A disentangled latent space for cross-site MRI harmonization* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2020), 720–729.

111. Zuo, L., Dewey, B. E., Carass, A., Liu, Y., He, Y., Calabresi, P. A. & Prince, J. L. *Information-Based Disentangled Representation Learning for Unsupervised MR Harmonization* in *International Conference on Information Processing in Medical Imaging* (2021), 346–359.

112. Studholme, C., Constable, R. T. & Duncan, J. S. Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model. *IEEE Transactions on Medical Imaging* **19,** 1115–1127 (2000).

113. Landman, B. A., Farrell, J. A., Patel, N., Mori, S. & Prince, J. L. *DTI fiber tracking: the importance of adjusting DTI gradient tables for motion correction. CATNAP-a tool to simplify and accelerate DTI analysis* in *Proc. Org Human Brain Mapping 13th Annual Meeting* (2007).

114. Alexander, D. C. & Barker, G. J. Optimal imaging parameters for fiber-orientation estimation in diffusion MRI. *NeuroImage* **27,** 357–367 (2005).

115. Cook, P., Bai, Y., Nedjati-Gilani, S., Seunarine, K., Hall, M., Parker, G. & Alexander, D. *Camino: Open-source diffusion-MRI reconstruction and processing* in *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine, Seattle, WA.* (2006), 2759.

116. Reinhold, J. C., Dewey, B. E., Carass, A. & Prince, J. L. *Evaluating the impact of intensity normalization on MR image synthesis* in *Medical Imaging 2019: Image Processing* (2019), 890–898.

117. Huo, Y., Xu, Z., Xiong, Y., Aboud, K., Parvathaneni, P., Bao, S., Bermudez, C., Resnick, S. M., Cutting, L. E. & Landman, B. A. 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage* **194,** 105–119 (2019).

118. Jia, J. *A package to compute segmentation metrics: seg-metrics* 2020.

119. Stone, M., Woo, J., Lee, J., Poole, T., Seagraves, A., Chung, M., Kim, E., Murano, E. Z., Prince, J. L. & Blemker, S. S. Structure and variability in human tongue muscle anatomy. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **6,** 499–507 (2018).

120. Takemoto, H. Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research* **44,** 95–107 (2001).

121. Kajee, Y., Pelteret, J.-P. & Reddy, B. D. The biomechanics of the human tongue. *International Journal for Numerical Methods in Biomedical Engineering* **29,** 492–514 (2013).

122. Kier, W. M. & Smith, K. K. Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society* **83,** 307–324 (1985).

123. Gomez, A. D., Elsaid, N., Stone, M. L., Zhuo, J. & Prince, J. L. Laplace-based modeling of fiber orientation in the tongue. *Biomechanics and Modeling in Mechanobiology* **17,** 1119–1130 (2018).

124. Gaige, T. A., Benner, T., Wang, R., Wedeen, V. J. & Gilbert, R. J. Three dimensional myoarchitecture of the human tongue determined in vivo by diffusion tensor imaging with tractography. *Journal of Magnetic Resonance Imaging* **26,** 654–661 (2007).

125. Shinagawa, H., Murano, E. Z., Zhuo, J., Landman, B., Gullapalli, R. P., Prince, J. L. & Stone, M. Tongue muscle fiber tracking during rest and tongue protrusion with oral appliances: A preliminary study with diffusion tensor imaging. *Acoustical Science and Technology* **29,** 291–294 (2008).

126. Alexander, A. L., Hasan, K. M., Lazar, M., Tsuruda, J. S. & Parker, D. L. Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **45,** 770–780 (2001).

127. Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W. & Wedeen, V. J. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine* **48,** 577–582 (2002).

128. Frank, L. R. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **47,** 1083–1099 (2002).

129. MacRobert, T. M. *Spherical harmonics: An elementary treatise on harmonic functions, with applications* (Dover publications, 1967).

130. Wikipedia contributors. *Spherical harmonics* 2022.

131. Alexander, D., Barker, G. & Arridge, S. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **48,** 331–340 (2002).

132. Tuch, D. S. Q-ball imaging. *Magnetic Resonance in Medicine* **52,** 1358–1372 (2004).

133. Tournier, J.-D., Calamante, F., Gadian, D. G. & Connelly, A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* **23,** 1176–1185 (2004).

134. Tournier, J.-D., Calamante, F. & Connelly, A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* **35,** 1459–1472 (2007).

135. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K. & Harel, N. Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magnetic resonance in medicine* **64,** 554–566 (2010).

136. Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34,** 144–155 (2007).

137. Lin, Z., Gong, T., Wang, K., Li, Z., He, H., Tong, Q., Yu, F. & Zhong, J. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. *Medical Physics* **46,** 3101–3116 (2019).

138. Wilkins, B., Lee, N., Gajawelli, N., Law, M. & Leporé, N. Fiber estimation and tractography in diffusion MRI: development of simulated brain images and comparison of multi-fiber analysis methods at clinical b-values. *NeuroImage* **109,** 341–356 (2015).

139. Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M. & Smith, S. M. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine* **50,** 1077–1088 (2003).

140. Wang, D. & Kleinberg, R. Analyzing quadratic unconstrained binary optimization problems via multicommodity flows. *Discrete Applied Mathematics* **157,** 3746–3753 (2009).

141. Booth, M., Reinhardt, S. & Roy, A. *Partitioning optimization problems for hybrid classcal/quantum execution* 2017.

142. Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M. & Nimmo-Smith, I. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in NeuroInformatics* **8,** 8 (2014).

143. Jenkinson, M., Beckmann, C., Behrens, T., Woolrich, M. & Smith, S. FSL. *NeuroImage* **62,** 782–90 (2012).

144. Schilling, K. G., Janve, V., Gao, Y., Stepniewska, I., Landman, B. A. & Anderson, A. W. Histological validation of diffusion MRI fiber orientation distributions and dispersion. *NeuroImage* **165,** 200–221 (2018).

145. Anderson, A. W. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine* **54,** 1194–1206 (2005).

146. Daducci, A., Canales-Rodrı, E. J., Descoteaux, M., Garyfallidis, E., Gur, Y., Lin, Y.-C., Mani, M., Merlet, S., Paquette, M., Ramirez-Manzanares, A., *et al.* Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. *IEEE Transactions on Medical Imaging* **33,** 384–399 (2013).

147. Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H. & Connelly, A. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202,** 116137 (2019).

148. Pierre, C. S., Dassonville, O., Chamorey, E., Poissonnet, G., Riss, J.-C., Ettaiche, M., Peyrade, F., Benezery, K., Chand, M.-E., Leyssalle, A., *et al.* Long-term functional outcomes and quality of life after oncologic surgery and microvascular reconstruction in patients with oral or oropharyngeal cancer. *Acta Oto-laryngologica* **134,** 1086–1093 (2014).

149. Buchaillard, S., Perrier, P. & Payan, Y. A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America* **126,** 2033–2051 (2009).

150. Sanguineti, V., Laboissiere, R. & Payan, Y. A control model of human tongue movements in speech. *Biological Cybernetics* **77,** 11–22 (1997).

151. Gomez, A. D., Stone, M. L., Woo, J., Xing, F. & Prince, J. L. Analysis of fiber strain in the human tongue during speech. *Computer Methods in Biomechanics and Biomedical Engineering* **23,** 312–322 (2020).

152. Xing, F., Stone, M., Goldsmith, T., Prince, J. L., El Fakhri, G. & Woo, J. Atlas-based tongue muscle correlation analysis from tagged and high-resolution magnetic resonance imaging. *Journal of Speech, Language, and Hearing Research* **62,** 2258–2269 (2019).

153. Pittman, L. J. & Bailey, E. F. Genioglossus and intrinsic electromyographic activities in impeded and unimpeded protrusion tasks. *Journal of NeuroPhysiology* **101,** 276–282 (2009).

154. Tankisi, H., Otto, M., Pugdahl, K. & Fuglsang-Frederiksen, A. Spontaneous electromyographic activity of the tongue in amyotrophic lateral sclerosis. *Muscle & Nerve* **48,** 296–298 (2013).

155. Sasaki, M., Onishi, K., Stefanov, D., Kamata, K., Nakayama, A., Yoshikawa, M. & Obinata, G. Tongue interface based on surface EMG signals of suprahyoid muscles. *Robomech Journal* **3,** 1–11 (2016).

156. Xing, F., Prince, J. L., Stone, M., Reese, T. G., Atassi, N., Wedeen, V. J., El Fakhri, G. & Woo, J. *Strain map of the tongue in normal and ALS speech patterns from tagged and diffusion MRI* in *Medical Imaging 2018: Image Processing* (2018), 248–256.

157. Rentschler, G. J. & Mann, M. B. The effects of glossectomy on intelligibility of speech and oral perceptual discrimination. *Journal of Oral Surgery (American Dental Association: 1965)* **38,** 348–354 (1980).

158. Chuanjun, C., Zhiyuan, Z., Shaopu, G., Xinquan, J. & Zhihong, Z. Speech after partial glossectomy: a comparison between reconstruction and nonreconstruction patients. *Journal of Oral and Maxillofacial Surgery* **60,** 404–407 (2002).

159. Bressmann, T., Sader, R., Whitehill, T. L. & Samman, N. Consonant intelligibility and tongue motility in patients with partial glossectomy. *Journal of Oral and Maxillofacial Surgery* **62,** 298–303 (2004).

160. Bressmann, T., Jacobs, H., Quintero, J. & Irish, J. C. Speech outcomes for partial glossectomy surgery: Measures of speech articulation and listener perception. *Head and Neck Cancer* **33,** 204 (2009).

161. Parthasarathy, V., Prince, J. L., Stone, M., Murano, E. Z. & NessAiver, M. Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *The Journal of the Acoustical Society of America* **121,** 491–504 (2007).

162. Zerhouni, E. A., Parish, D. M., Rogers, W. J., Yang, A. & Shapiro, E. P. Human heart: tagging with MR imaging—a method for noninvasive assessment of myocardial motion. *Radiology* **169,** 59–63 (1988).

163. Liu, X., Abd-Elmoniem, K. Z., Stone, M., Murano, E. Z., Zhuo, J., Gullapalli, R. P. & Prince, J. L. Incompressible deformation estimation algorithm (IDEA) from tagged MR images. *IEEE Transactions on Medical Imaging* **31,** 326–340 (2011).

164. Xing, F., Liu, X., Reese, T., Stone, M., Wedeen, V., Prince, J. L., El Fakhri, G. & Woo, J. Muscle strain analysis from diffusion tractography and dynamic magnetic resonance imaging of the moving tongue. *The Journal of the Acoustical Society of America* **150,** A190–A190 (2021).

165. Elsaid, N. M., Prince, J. L., Roys, S., Gullapalli, R. P. & Zhuo, J. Phase image texture analysis for motion detection in diffusion MRI (PITA-MDD). *Magnetic Resonance Imaging* **62,** 228–241 (2019).

166. Liang, X., Su, P., Patil, S. G., Elsaid, N. M., Roys, S., Stone, M., Gullapalli, R. P., Prince, J. L. & Zhuo, J. Prospective motion detection and re-acquisition in diffusion MRI using a phase image-based method—Application to brain and tongue imaging. *Magnetic Resonance in Medicine* **86,** 725–737 (2021).

167. Fischer, S. E., McKinnon, G. C., Maier, S. E. & Boesiger, P. Improved myocardial tagging contrast. *Magnetic Resonance in Medicine* **30,** 191–200 (1993).

168. Woo, J., Murano, E. Z., Stone, M. & Prince, J. L. Reconstruction of High-Resolution Tongue Volumes From MRI. en. *IEEE Transactions on Biomedical Engineering* **59,** 3511–3524 (2012).

169. Woo, J., Lee, J., Murano, E. Z., Xing, F., Al-Talib, M., Stone, M. & Prince, J. L. A high-resolution atlas and statistical model of the vocal tract from structural MRI. en. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **3,** 47–60 (2015).

170. Miyawaki, K. A study of the muscular of the human tongue. *Ann. Bull. RILP, Univ. Tokyo* **8,** 23–50 (1974).

171. Nie, D., Gao, Y., Wang, L. & Shen, D. *ASDNet: attention based semi-supervised deep networks for medical image segmentation* in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018), 370–378.

172. Sinha, A. & Dolz, J. Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics* **25,** 121–130 (2020).

173. Tournier, J. D., Calamante, F., Connelly, A., *et al. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions* in *Proceedings of the International Society for Magnetic Resonance in Medicine* (2010).

174. Tourdias, T., Saranathan, M., Levesque, I. R., Su, J. & Rutt, B. K. Visualization of Intra-Thalamic Anatomy with White-Matter-nulled MPRAGE at 7T. *International Society for Magnetic Resonance in Medicine* (2014).

175. Zhuo, J., Roys, S., Hebel, J., Owens, E., Raghavan, P., Gandhi, D. & Gullapalli, R. P. Visualization of the VIM Thalamic Nucleus using Synthesized MPRAGE Images. *International Society for Magnetic Resonance in Medicine* (2017).

176. Liu, X., Xing, F., Prince, J. L., Stone, M., El Fakhri, G. & Woo, J. *Structure-aware unsupervised tagged-to-cine MRI synthesis with self disentanglement* in *Medical Imaging 2022: Image Processing* (2022), 470–476.

Vita

Muhan Shao received her Bachelor degree in Electrical Engineering from Zhejiang University, Zhejiang, China in 2016. In 2016, she started her Ph.D. in the department of Electrical and Computer Engineering from the Johns Hopkins University and joined the Image Analysis and Communications Laboratory under the supervision of Dr. Jerry L. Prince. Her research interests include computer vision, deep learning, and medical image analysis. During her time as a graduate student, she has collaborated on projects studying normal pressure hydrocephalus, mild traumatic brain injuries, and oral cancer.